

A Comparison Between the Performance of Wayback Machines

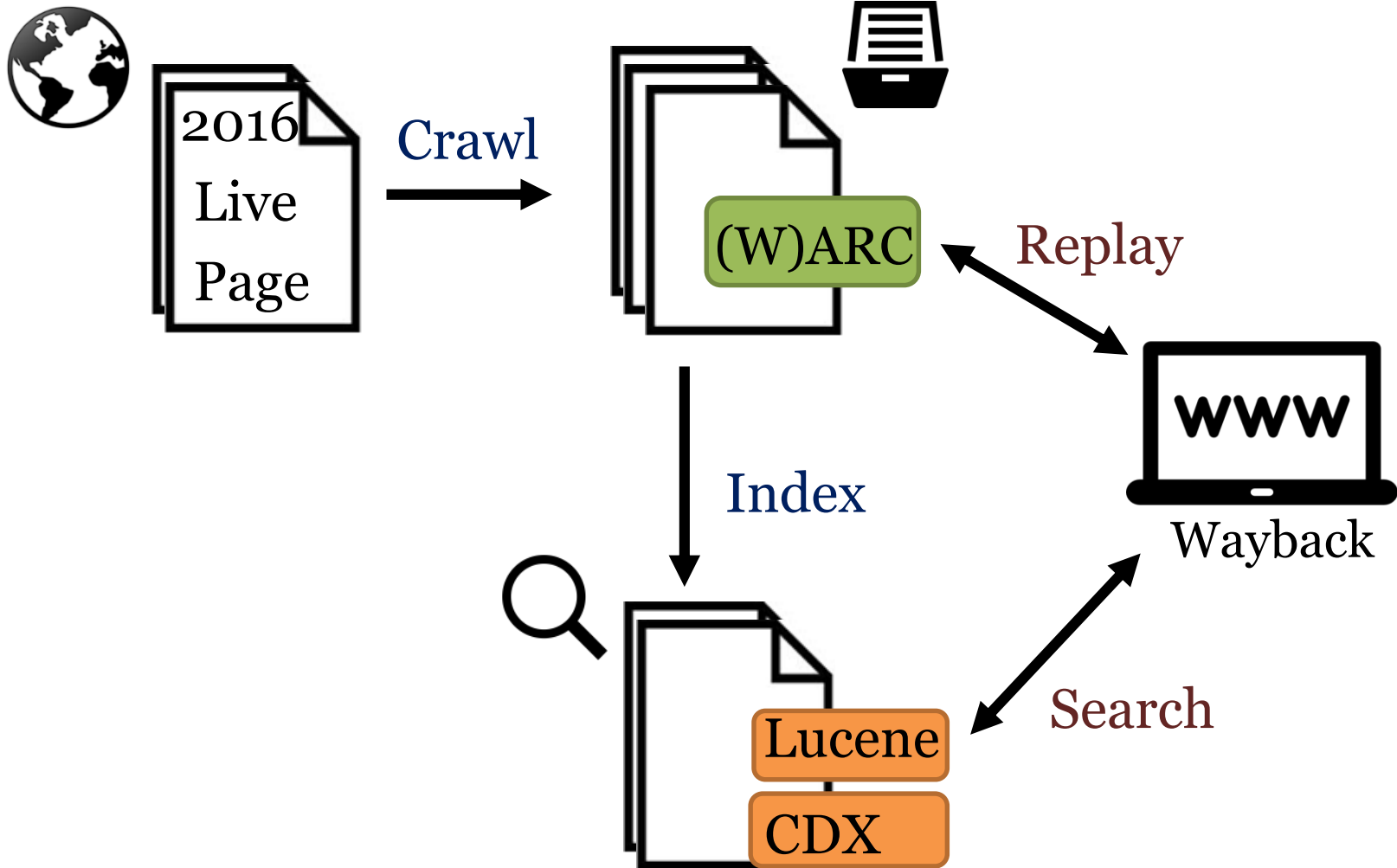
Fernando Melo fernando.melo@fccn.pt

Main reasons for this study

Outdated Wayback

Evaluate possible alternatives

How does a Web archive work?



What is a Wayback Machine?

What is a Wayback Machine?

Software Component

Replay Archived Web Pages

Search by URL and Date

What is a Wayback Machine?



between: and:

[Advanced search](#)

Did you want to see webpages with the text: <http://edition.cnn.com/>?

Versions of the archived web pages

We archived 21 versions of the Web page <http://edition.cnn.com/> from 1 January, 1996 and 8 April, 2016.

2001 0	2002 0	2003 0	2004 0	2005 0	2006 0	2007 0	2008 5	2009 5	2010 2	2011 3	2012 1	2013 1	2014 4	2015 0
							15 Feb	20 May	29 May	22 Jan	23 Jan	6 Nov	6 Sep	
							14 Mar	26 Jun	4 Aug	22 May			23 Nov	
							14 Mar	30 Sep		2 Jul			23 Nov	
							22 Oct	30 Sep					25 Nov	
							22 Oct	22 Dec						

What is a Wayback Machine?

EDITION: INTERNATIONAL | U.S. | MÉXICO | ARABIC
Sign up | Log in



SEARCH

Home
Video
World
U.S.
Africa
Asia
Europe
Latin America
Middle East
Business
World Sport
Entertainment
Tech
Travel
iReport

July 2, 2011 – Updated 1544 GMT (2344 HKT)




Kvitová stuns Sharapova to claim Wimbledon crown

Petra Kvitová stuns Maria Sharapova to claim her first Wimbledon women's singles title. The 21-year-old Czech had gone into the final as the underdog against Sharapova. [FULL STORY](#)

Nadal to play Djokovic in men's final
Williams sisters exit Wimbledon


Unrest in the Arab World
Libyan rebels wounded in firefight
Syrian president sacks governor
Bahrain govt and opposition to meet
At least 9 dead in Syrian protests




Monaco wedding celebrations continue

Monaco's Prince Albert and his South African bride Charlene Wittstock reaffirm their union in a religious ceremony at the prince's palace. [FULL STORY](#) | [A TROUBLED ROYAL HOUSEHOLD](#)


Highlights



The American soccer star in Palestine



The royals' Canada day outing



Janet Jackson's duet with brother Michael

[Audio updates](#) | [Make CNN Your Homepage](#)

ADVERTISEMENT

Hi! Log in or sign up to personalize!

POPULAR ON FACEBOOK

MOST POPULAR

WEATHER

MARKETS

Common Wayback Machine Issues

Slow Replay




LOADING

Not Found Errors

404

Page not found

Not Found Errors

 [Arquivo da Web Portuguesa](#) - ligações exteriores, formulários e caixas de pesquisa poderão não funcionar corretamente. URL: <http://ec.europa.eu/> Data: 13:16:11 5 Setembro, 2014 [[lecionar](#)]

[A-Z Index](#) | [Archives](#) | [Sitemap](#) | [About this site](#) | [Legal notice](#) | [Cookies](#) | [Contact](#) | [Search](#)



EUROPEAN COMMISSION

European Commission


- [BG](#) Добре дошли в Европейската комисия
- [CS](#) Vítejte v Evropské komisi
- [DA](#) Velkommen til Europa-Kommissionen
- [DE](#) Willkommen bei der Europäischen Kommission
- [ET](#) Tere tulemast Euroopa Komisjoni
- [EL](#) Καλώς ήρθατε στην Ευρωπαϊκή Επιτροπή
- [EN](#) Welcome to the European Commission
- [ES](#) Bienvenido a la Comisión Europea
- [FR](#) Bienvenue à la Commission européenne
- [HR](#) Dobro došli u Europsku komisiju
- [IT](#) Benvenuti alla Commissione europea
- [LV](#) Laipni lūdzam Eiropas Komisijā
- [LT](#) Sveiki atvykę į Europos Komisiją

 [Fájlte go dtí an Coimisiún Eorpach!](#)

- [HU](#) Üdvözljük az Európai Bizottságnál
- [MT](#) Merħba fil-Kummissjoni Ewropea
- [NL](#) Welkom bij de Europese Commissie
- [PL](#) Witamy na stronach Komisji Europejskiej
- [PT](#) Bem-vindos à Comissão Europeia
- [RO](#) Bun venit la Comisia Europeană
- [SK](#) Vitajte v Európskej komisii
- [SL](#) Dobrodošli na Evropski komisiji
- [FI](#) Tervetuloa Euroopan komission

 [Välkommen till Europeiska kommissionen](#)

Not Found Errors

 [Arquivo da Web Portuguesa](#) - ligações exteriores, formulários e caixas de pesquisa poderão não funcionar corretamente. URL: <http://ec.europa.eu/> Data: 13:16:11 5 Setembro, 2014 [[esconder](#)]

[A-Z Index](#) | [Archives](#) | [Sitemap](#) | [About this site](#) | [Legal notice](#) | [Cookies](#) | [Contact](#) | [Search](#)

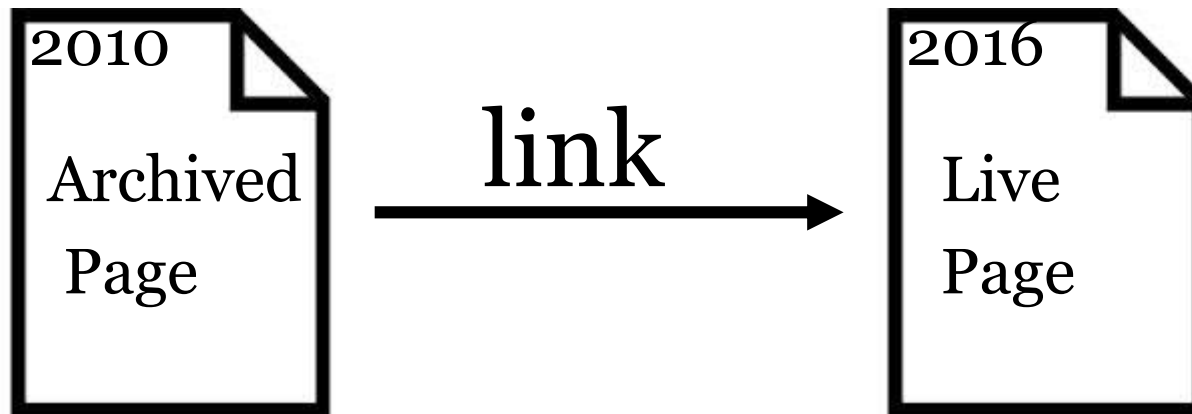


EUROPEAN COMMISSION

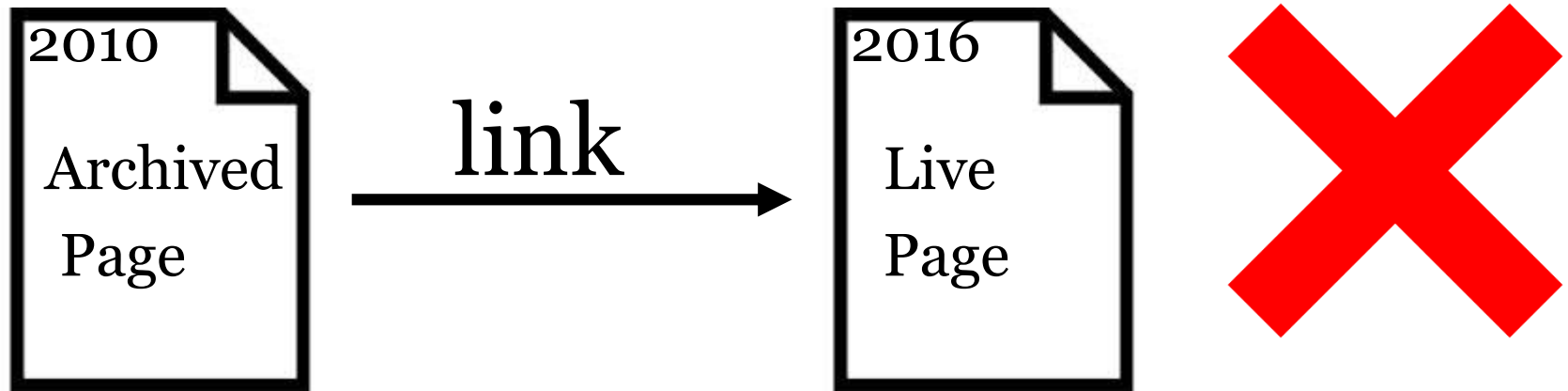
European Commission

- [BG](#) Добре дошли в Европейската комисия
- [CS](#) Vítejte v Evropské komisi
- [DA](#) Velkommen til Europa-Kommissionen
- [DE](#) Willkommen bei der Europäischen Kommission
- [ET](#) Tere tulemast Euroopa Komisjoni
- [EL](#) Καλώς ήρθατε στην Ευρωπαϊκή Επιτροπή
- [EN](#) Welcome to the European Commission
- [ES](#) Bienvenido a la Comisión Europea
- [FR](#) Bienvenue à la Commission européenne
- [HR](#) Dobro došli u Europsku komisiju
- [IT](#) Benvenuti alla Commissione europea
- [LV](#) Laipni lūdzam Eiropas Komisijā
- [LT](#) Sveiki atvykę į Europos Komisiją
- [GA](#) Fáilte go dtí an Coimisiún Eorpach
- [HU](#) Üdvözljük az Európai Bizottságnál
- [MT](#) Merħba fil-Kummissjoni Ewropea
- [NL](#) Welkom bij de Europese Commissie
- [PL](#) Witamy na stronach Komisji Europejskiej
- [PT](#) Bem-vindos à Comissão Europeia
- [RO](#) Bun venit la Comisia Europeană
- [SK](#) Vitajte v Európskej komisii
- [SL](#) Dobrodošli na Evropski komisiji
- [FI](#) Tervetuloa Euroopan komissioon
- [SV](#) Välkommen till Europeiska kommissionen

Live-Web Leaks



Live-Web Leaks

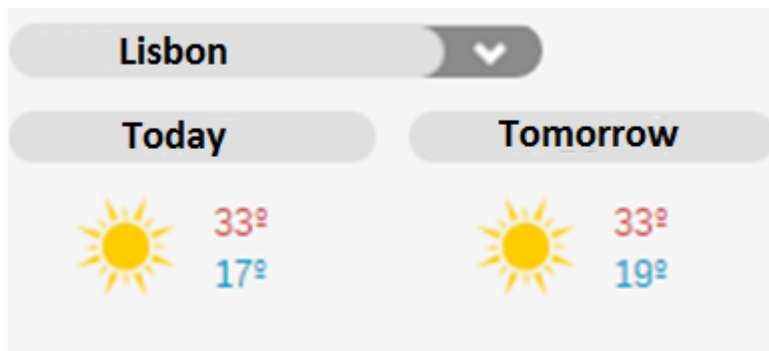


Live-Web Leaks

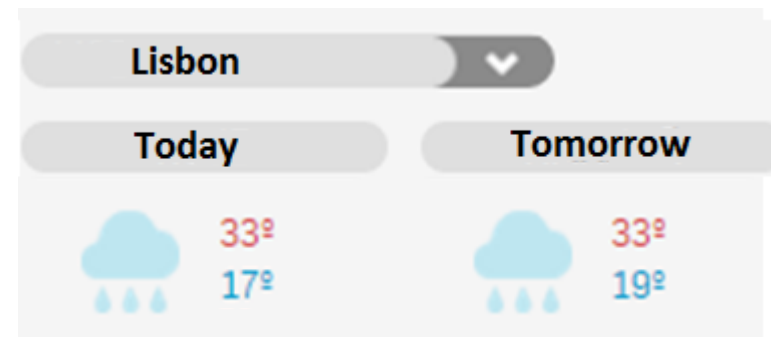


Live-Web Leaks

Original Web Page
July 14th, 2012

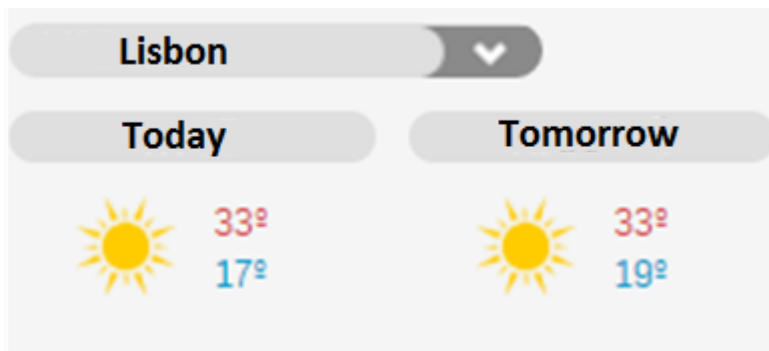


Archived Web Page
July 14th, 2012

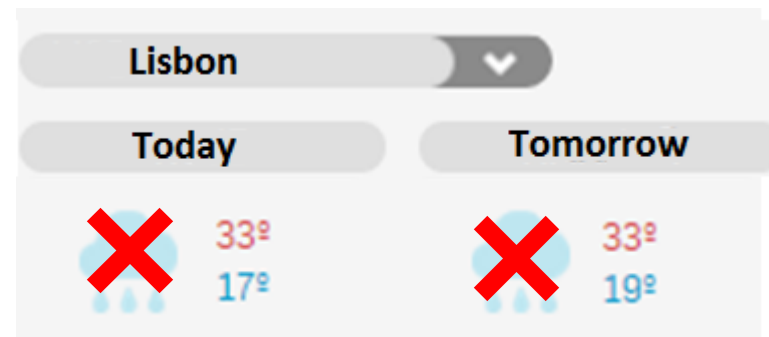


Live-Web Leaks

Original Web Page
July 14th, 2012

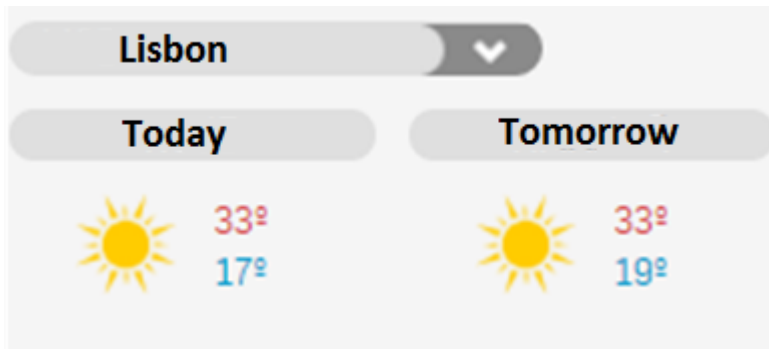


Archived Web Page
July 14th, 2012

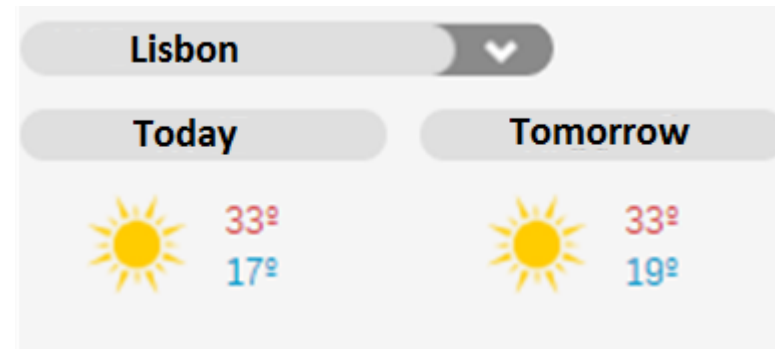


Live-Web Leaks

Original Web Page
July 14th, 2012



Archived Web Page
July 14th, 2012



Let's evaluate the **performance** of
Wayback Machine Software!

Wayback Machines

Arquivo.pt Wayback

OpenWayback

PyWb

Wayback Machines

Arquivo.pt Wayback

Derives from version 1.2.1 of Open Source
Wayback Machine (2008)

Java

Used by Arquivo.pt

Outdated - Presents several replay issues

PyWb Wayback

Developed by Ilya Kreymer

Python

Used by

<http://rhizome.org>

<http://webrecorder.io>

<http://perma.cc>

OpenWayback

Released by the Internet Archive

Maintained by the IIPC

Java

OpenWayback - Users

National and University Library of Iceland

The British Library

Archive-It Mirror @ ODU

Stanford Web Archive Portal

The Library of Congress

Bibliotheca Alexandrina

York University Digital Library

Bibliothèque nationale de France

University of North Texas Libraries

The .EU Collection - 2014

The .EU Collection - 2014

Domains can be sold to anyone with a valid address in the European Union

European Institutions, Online Shops, and Web Spam

250 million documents from 34 thousand seeds

6TB

Methodology

400 URLs from the .EU

WebPageTest service

4 Wayback Configurations

HAR – to record performance data

Methodology

[HOME](#) [TEST HISTORY](#) [FORUMS](#) [DOCUMENTATION](#) [ABOUT](#)

Test a website's performance

[Analytical Review](#) [Visual Comparison](#) [Traceroute](#)

Test Location [Select from Map](#)

Browser

Advanced Settings ▼

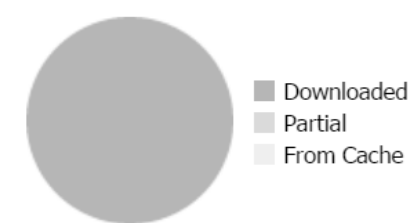
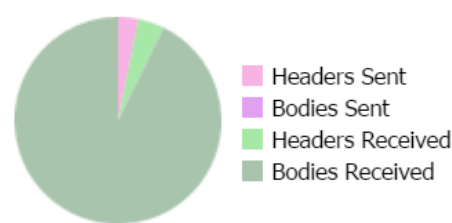
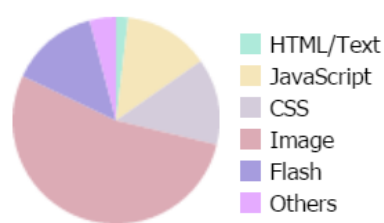
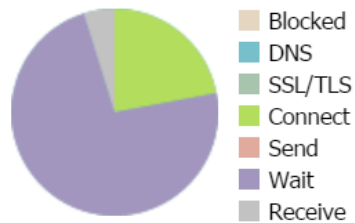
Test Settings [Advanced](#) [Chrome](#) [Auth](#) [Script](#) [Block](#) [SPOF](#) [Custom](#)

Connection

START TEST

Methodology

Page Load: 55.79s, 174 Requests 10/28/2015, 11:23:08 AM Run 1, First View for http://p27.arquivo.pt:8282/replay/20141121205453/lu.overnightprints.eu



Run 1, First View for http://p27.arquivo.pt:8282/replay/20141121205453/lu.overnightprints.eu

Request	Status	Size	Time
GET lu.overnightpr	302	0	2.81s
GET lu.overnightpr	200	46.2 Ki	3.07s
GET wombat.js	200	68.5 Ki	3.02s
GET wb.css	200	1 KB	3.11s
GET reset.css?stc=	302	0	3.23s
GET general.css?st	302	0	3.16s
GET formate.css?sl	302	0	3.17s
GET css_de.css?stc	302	0	3.3s
GET styles.css?stc:	302	0	3.44s
GET productpage.c	302	0	3.59s
GET anythingslider	302	0	3.6s
GET anythingslider	302	0	6.58s
GET anythingslider	302	0	3.83s

Methodology

Only test each URL once

Tested using WebPageTest public servers

Response timeout of 2 minutes

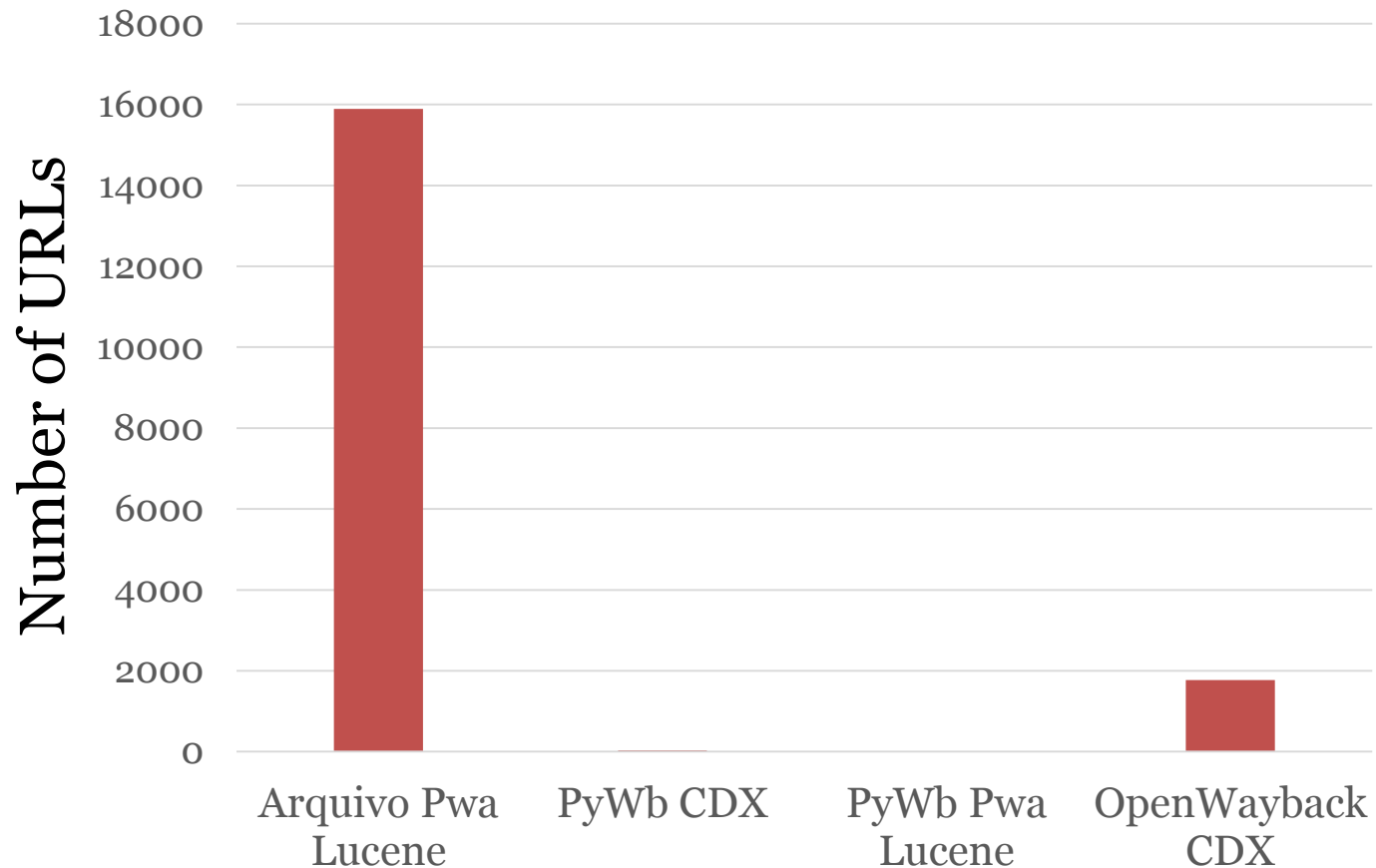
Error Code – Leak to the live Web

Wayback Specifications

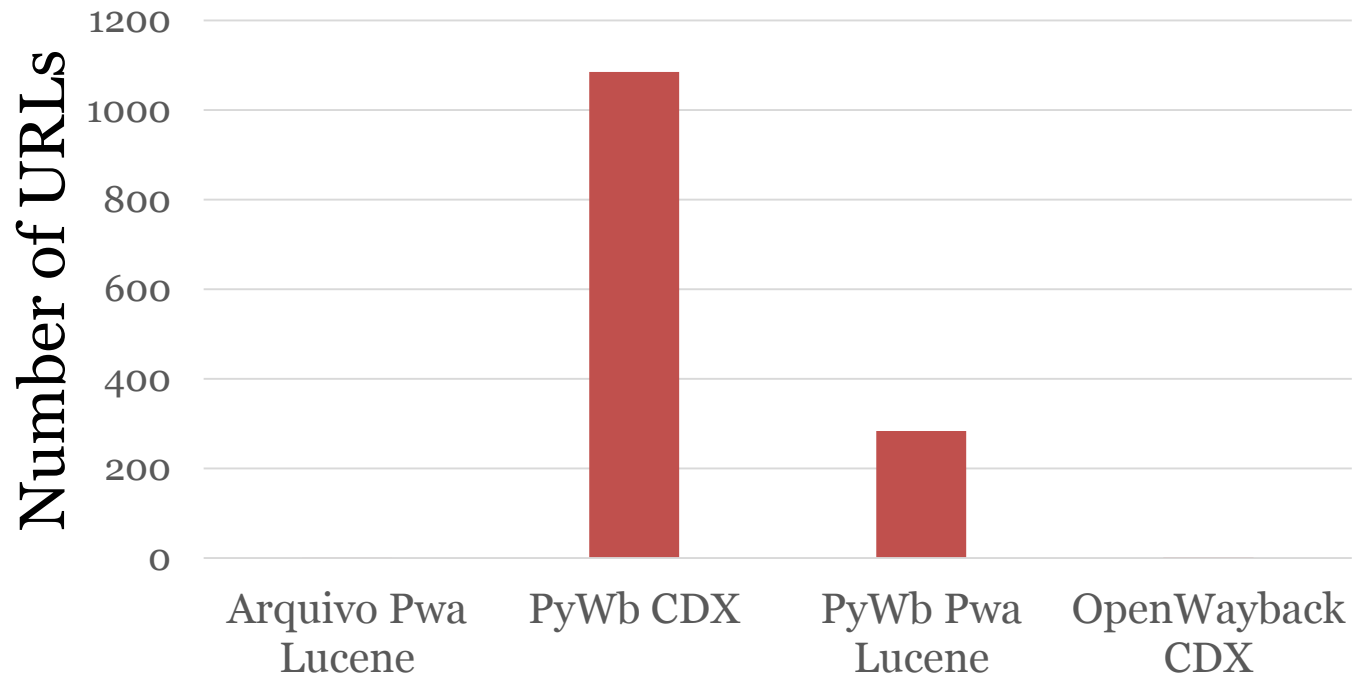
Wayback	Year
Arquivo Pwa Lucene	2008
PyWb CDX	2015
PyWb Pwa Lucene	2015
OpenWayback CDX	2015

Replay Quality – HTTP Status and Error Codes

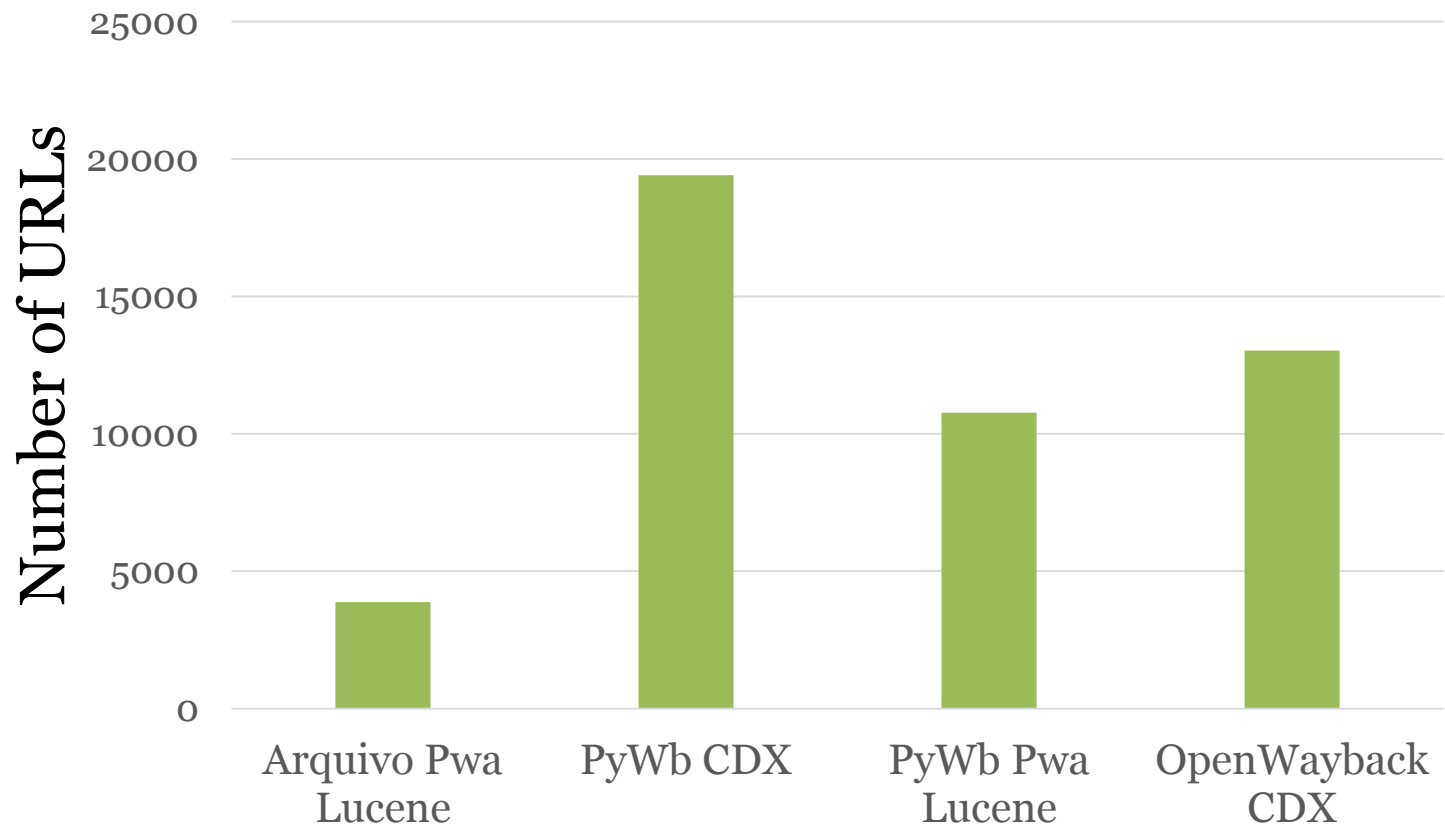
Results – Live Web Leaks



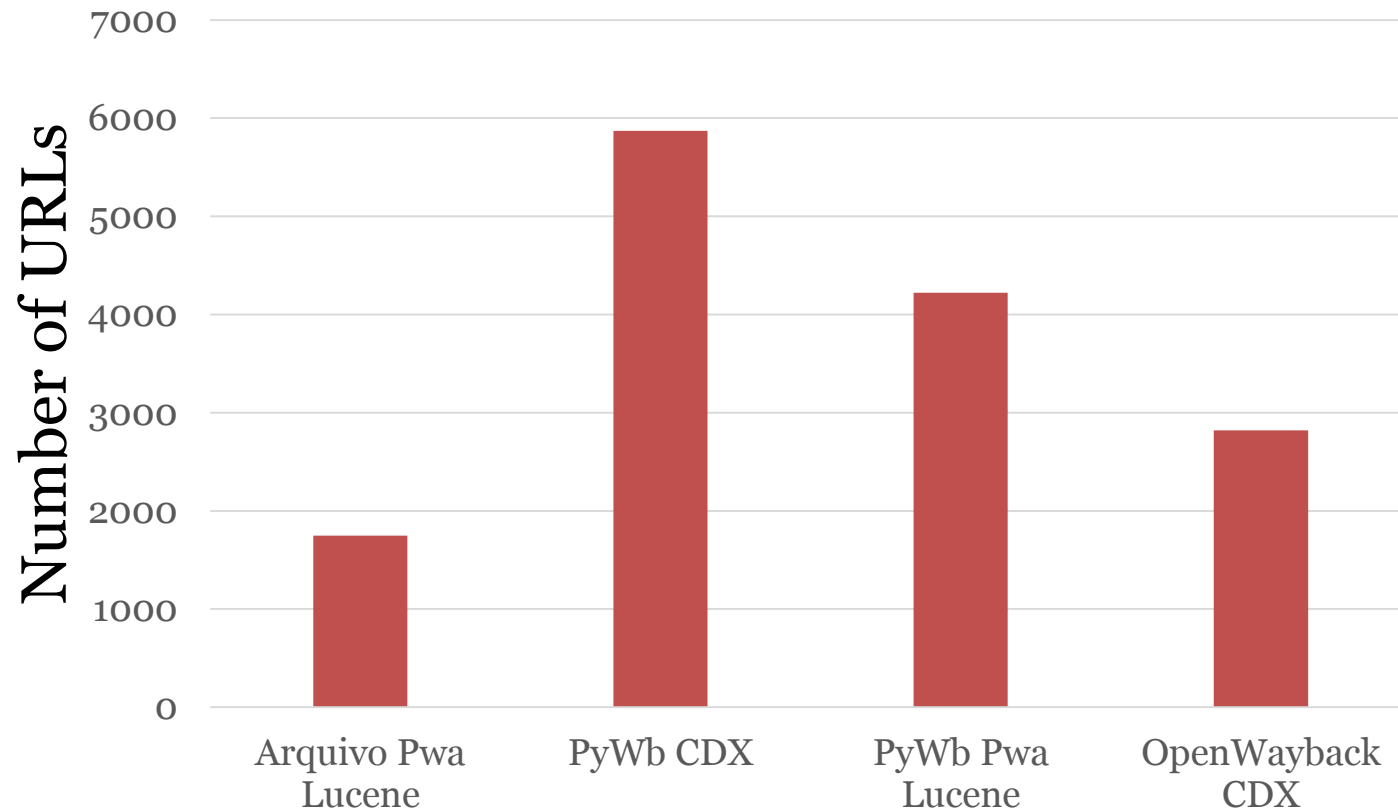
Results – Timeout Error



Results – 200 OK Status Code



Results – 404 Error HTTP Code

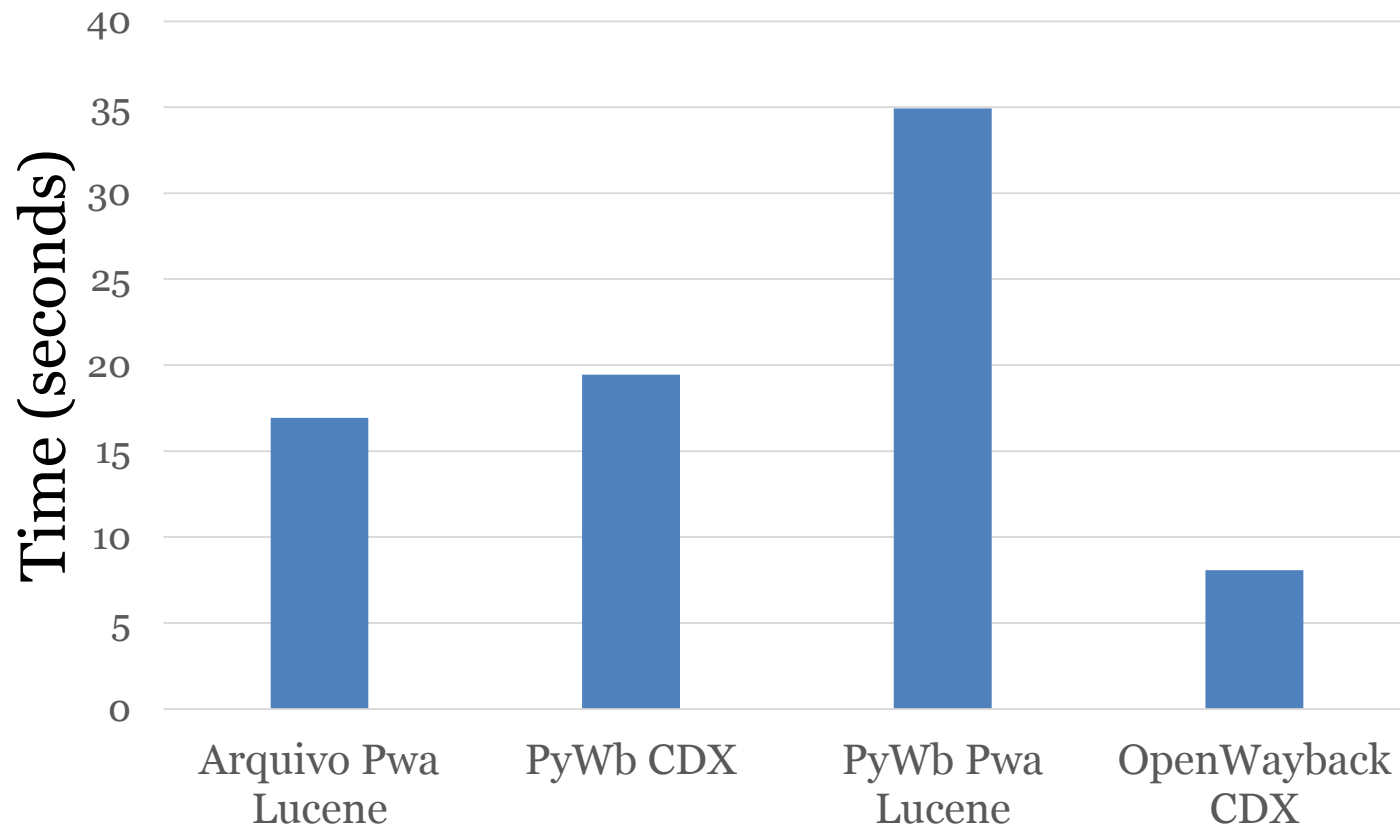


Results – Summary Table

Wayback	Success	Error	Success/ Error
Arquivo	3 930	17 711	0.22
PyWb CDX	19 415	7 082	2.74
PyWb Pwa Lucene	11 087	4 652	2.38
OpenWayback	13 068	4 668	2.80

Response Speed

Results – Average Load Time



Conclusions

PyWb presented the biggest number of 200 OK HTTP status codes

OpenWayback was the fastest Wayback

Replace or Update Arquivo.pt's Wayback!

Future Work

Test with older collections to evaluate the performance of Wayback Machine software

Test with private instance of WebpageTest server to be able to execute more tests and to control the server workload

References

<https://github.com/Fernando-Melo/WaybackComparison>