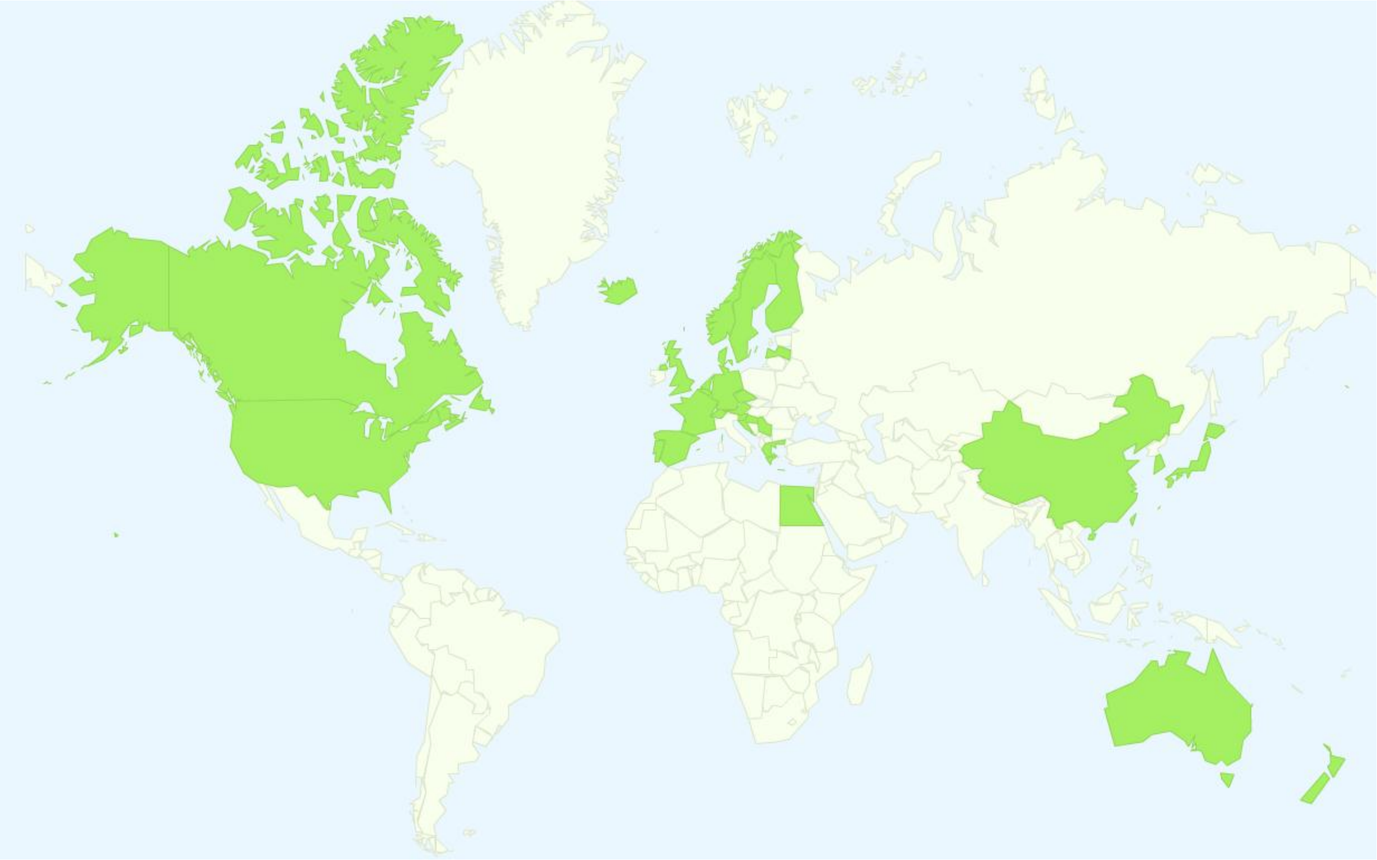
Creating a billion-scale searchable web archive

Daniel Gomes, Miguel Costa, David Cruz,
João Miranda and Simão Fontes

PORTUGUESE
WEB ARCHIVE

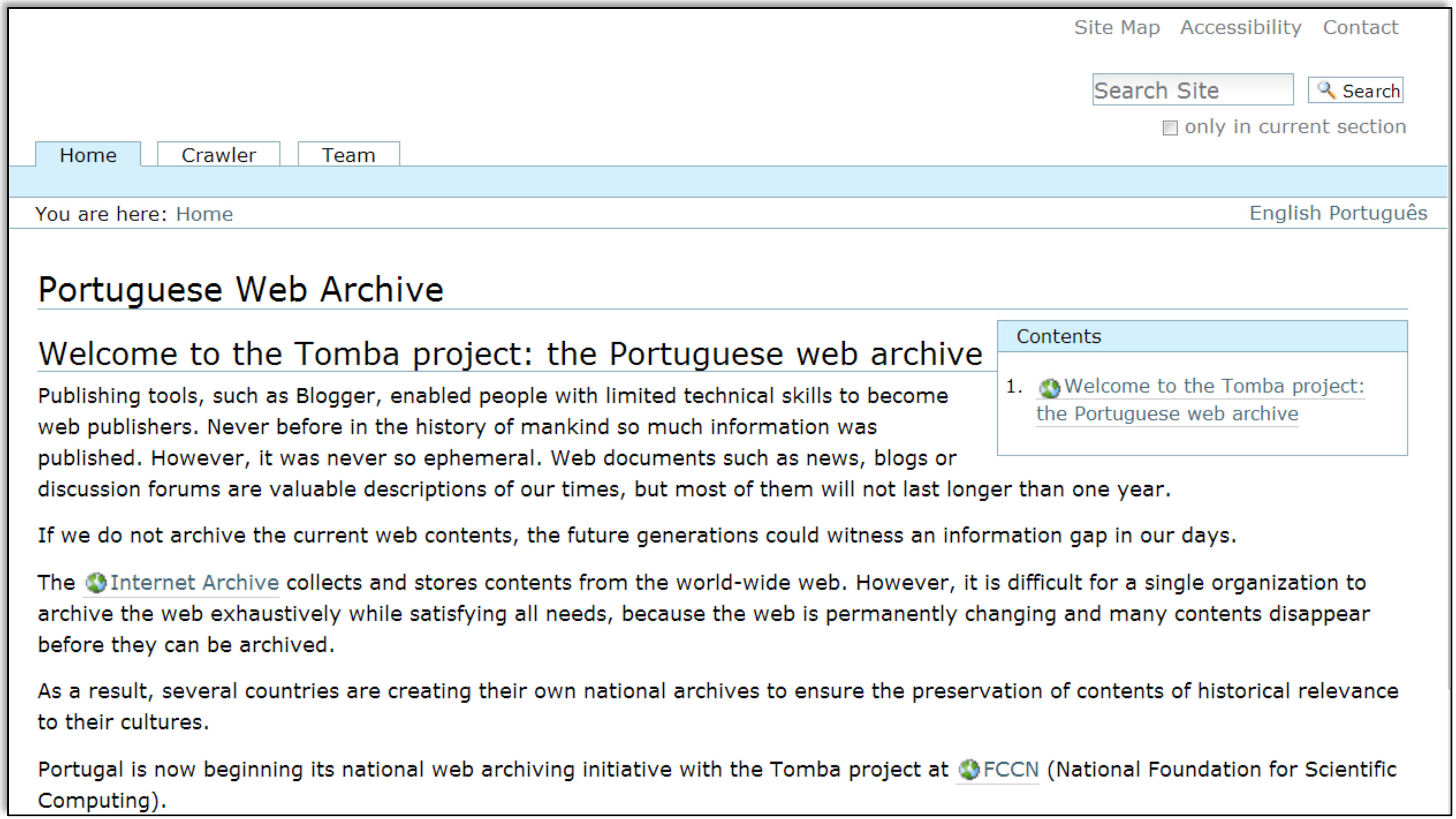
Web archiving initiatives are spreading around the world



- At least 6.6 PB were archived since 1996

The Portuguese Web Archive aims to preserve Portuguese cultural heritage

The Portuguese Web Archive project started in 2008



Portuguese Web Archive

Welcome to the Tomba project: the Portuguese web archive

Contents

1. Welcome to the Tomba project: the Portuguese web archive

Publishing tools, such as Blogger, enabled people with limited technical skills to become web publishers. Never before in the history of mankind so much information was published. However, it was never so ephemeral. Web documents such as news, blogs or discussion forums are valuable descriptions of our times, but most of them will not last longer than one year.

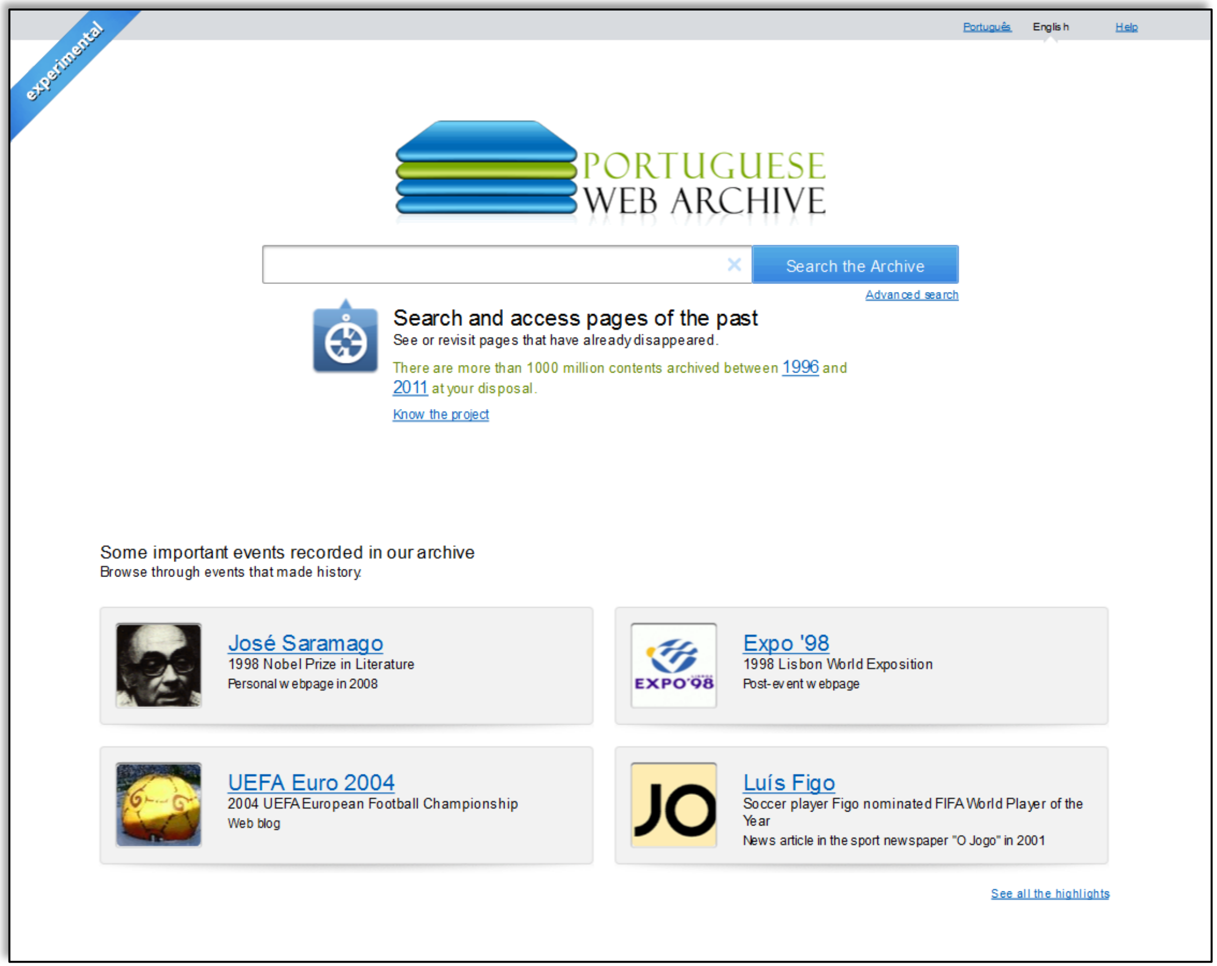
If we do not archive the current web contents, the future generations could witness an information gap in our days.

The Internet Archive collects and stores contents from the world-wide web. However, it is difficult for a single organization to archive the web exhaustively while satisfying all needs, because the web is permanently changing and many contents disappear before they can be archived.

As a result, several countries are creating their own national archives to ensure the preservation of contents of historical relevance to their cultures.

Portugal is now beginning its national web archiving initiative with the Tomba project at FCCN (National Foundation for Scientific Computing).

It was announced last year (2012)



Search and access pages of the past

See or revisit pages that have already disappeared.

There are more than 1000 million contents archived between 1996 and 2011 at your disposal.

Know the project

Some important events recorded in our archive

Browse through events that made history.

José Saramago
1998 Nobel Prize in Literature
Personal webpage in 2008

Expo '98
1998 Lisbon World Exposition
Post-event webpage

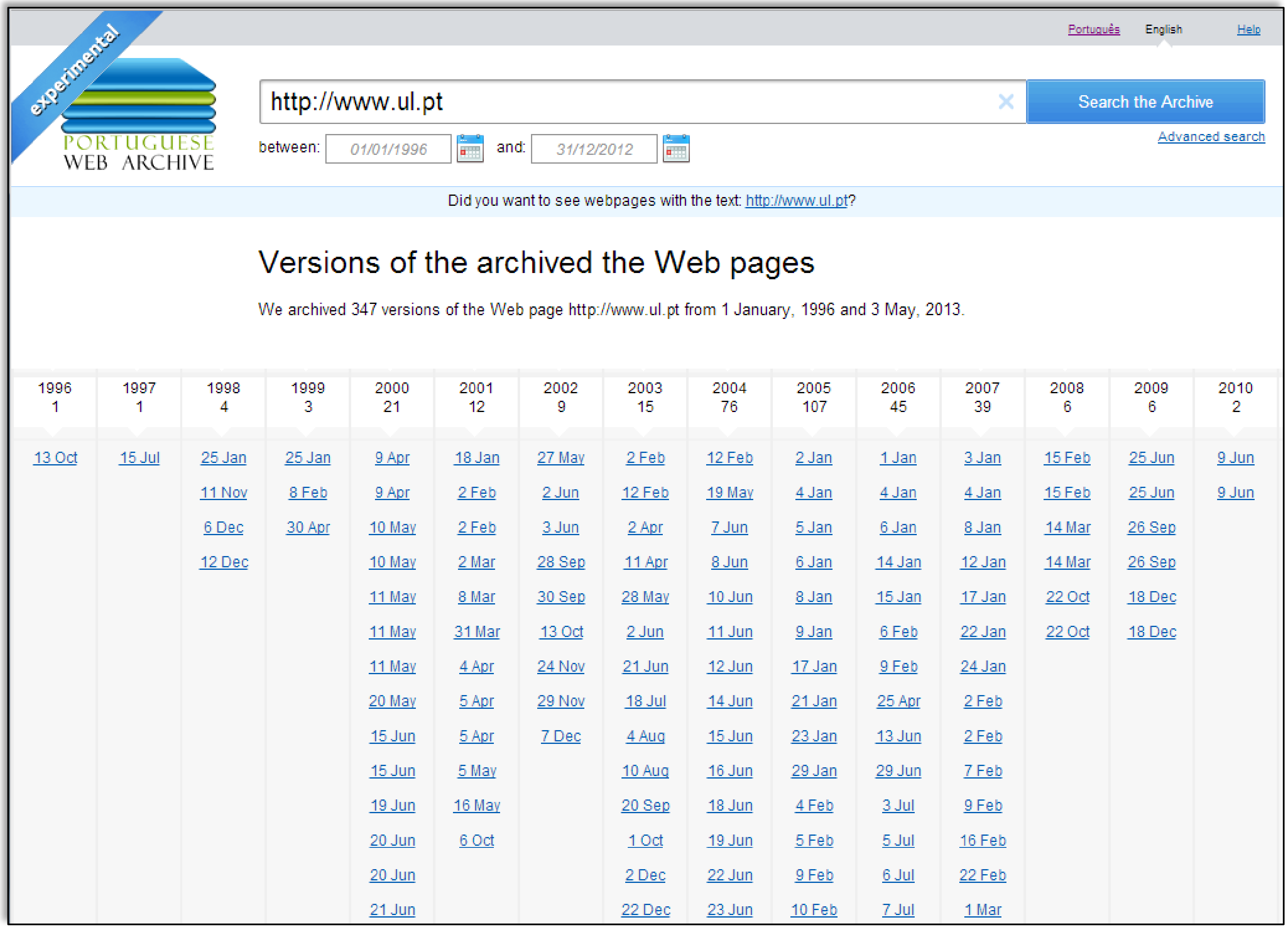
UEFA Euro 2004
2004 UEFA European Football Championship
Web blog

Luís Figo
Soccer player Figo nominated FIFA World Player of the Year
News article in the sport newspaper "O Jogo" in 2001

See all the highlights

- Public and free at archive.pt

Provides **version history** like the Internet Archive Wayback Machine



http://www.ul.pt

between: 01/01/1996 and: 31/12/2012

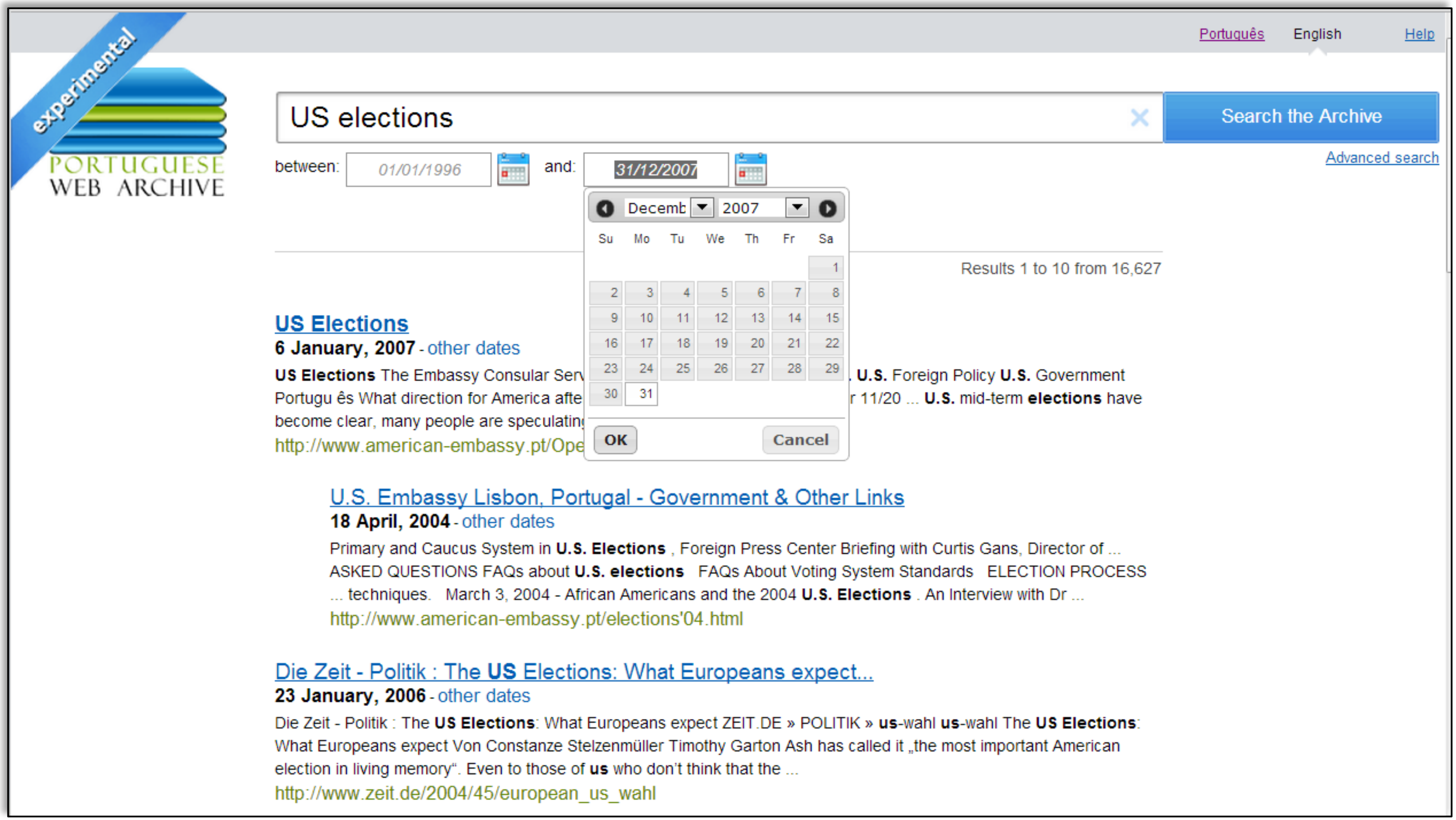
Did you want to see webpages with the text: http://www.ul.pt?

Versions of the archived the Web pages

We archived 347 versions of the Web page http://www.ul.pt from 1 January, 1996 and 3 May, 2013.

1996 1	1997 1	1998 4	1999 3	2000 21	2001 12	2002 9	2003 15	2004 76	2005 107	2006 45	2007 39	2008 6	2009 6	2010 2
13 Oct	15 Jul	25 Jan	25 Jan	9 Apr	18 Jan	27 May	2 Feb	12 Feb	2 Jan	1 Jan	3 Jan	15 Feb	25 Jun	9 Jun
		11 Nov	8 Feb	9 Apr	2 Feb	2 Jun	12 Feb	19 May	4 Jan	4 Jan	4 Jan	15 Feb	25 Jun	9 Jun
		6 Dec	30 Apr	10 May	2 Feb	3 Jun	2 Apr	7 Jun	5 Jan	6 Jan	8 Jan	14 Mar	26 Sep	
		12 Dec		10 May	2 Mar	28 Sep	11 Apr	8 Jun	6 Jan	14 Jan	12 Jan	14 Mar	26 Sep	
				11 May	8 Mar	30 Sep	28 May	10 Jun	8 Jan	15 Jan	17 Jan	22 Oct	18 Dec	
				11 May	31 Mar	13 Oct	2 Jun	11 Jun	9 Jan	6 Feb	22 Jan	22 Oct	18 Dec	
				11 May	4 Apr	24 Nov	21 Jun	12 Jun	17 Jan	9 Feb	24 Jan			
				20 May	5 Apr	29 Nov	18 Jul	14 Jun	21 Jan	25 Apr	2 Feb			
				15 Jun	5 Apr	7 Dec	4 Aug	15 Jun	23 Jan	13 Jun	2 Feb			
				15 Jun	5 May		10 Aug	16 Jun	29 Jan	29 Jun	7 Feb			
				19 Jun	16 May		20 Sep	18 Jun	4 Feb	3 Jul	9 Feb			
				20 Jun	6 Oct		1 Oct	19 Jun	5 Feb	5 Jul	16 Feb			
				20 Jun			2 Dec	22 Jun	9 Feb	6 Jul	22 Feb			
				21 Jun			22 Dec	23 Jun	10 Feb	7 Jul	1 Mar			

But also **full-text search** over 1.2 billion web files archived since 1996



US elections

Results 1 to 10 from 16,627

US Elections
6 January, 2007 - other dates
US Elections The Embassy Consular Serv . **U.S.** Foreign Policy **U.S.** Government
Portugu ês What direction for America afte r 11/20 ... **U.S.** mid-term **elections** have
become clear, many people are speculatin
http://www.american-embassy.pt/Ope

U.S. Embassy Lisbon, Portugal - Government & Other Links
18 April, 2004 - other dates
Primary and Caucus System in **U.S. Elections** , Foreign Press Center Briefing with Curtis Gans, Director of ... ASKED QUESTIONS FAQs about **U.S. elections** FAQs About Voting System Standards ELECTION PROCESS ... techniques. March 3, 2004 - African Americans and the 2004 **U.S. Elections** . An Interview with Dr ...
http://www.american-embassy.pt/elections'04.html

Die Zeit - Politik : The **US** Elections: What Europeans expect...
23 January, 2006 - other dates
Die Zeit - Politik : The **US Elections**: What Europeans expect ZEIT.DE » POLITIK » **us**-wahl **us**-wahl The **US Elections**: What Europeans expect Von Constanze Stelzenmüller Timothy Garton Ash has called it „the most important American election in living memory“. Even to those of **us** who don't think that the ...
http://www.zeit.de/2004/45/european_us_wahl

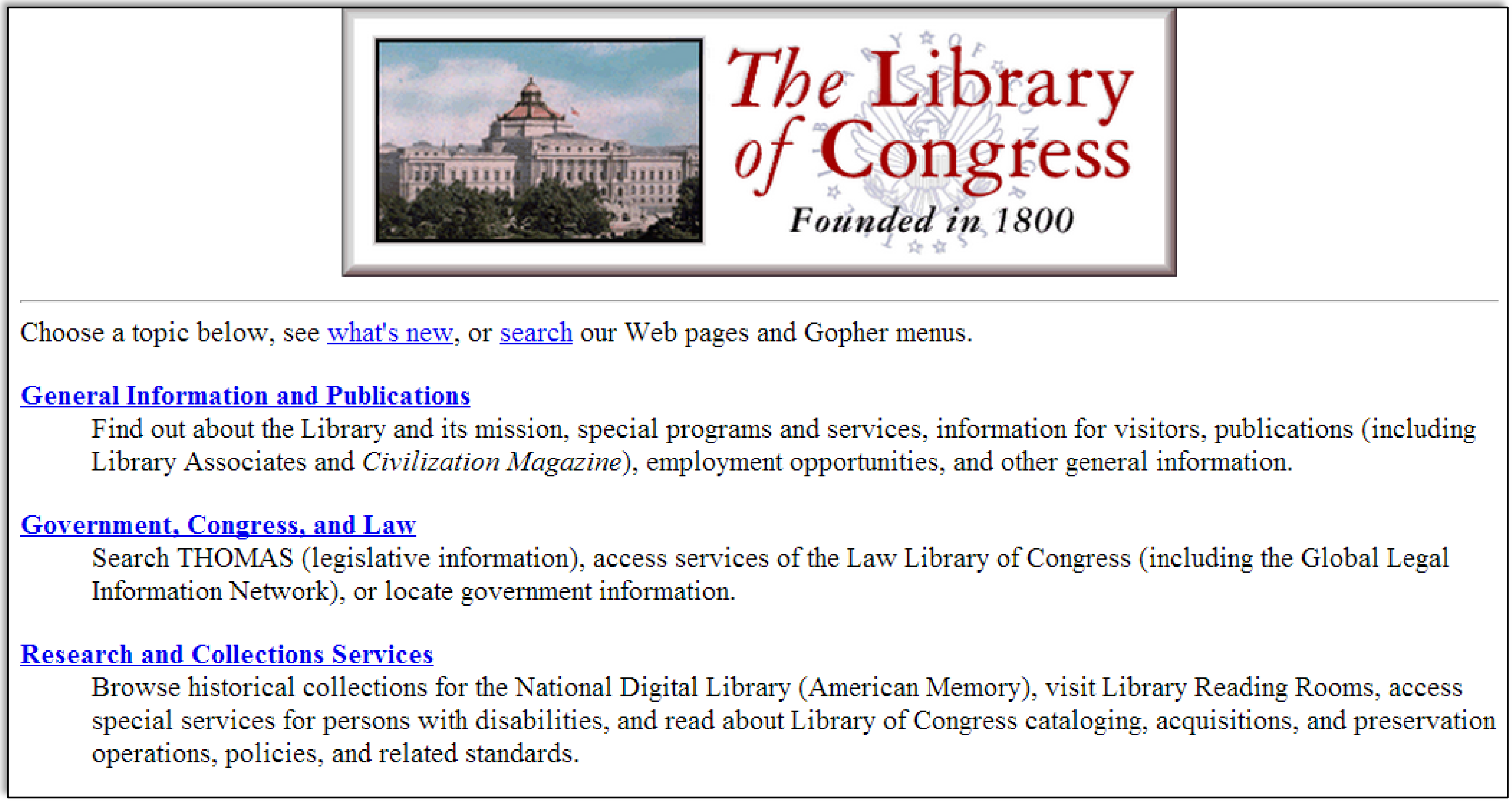
Now…the details.

Acquiring web data

Integration of third-party collections archived before 2007

- Integration of historical collections (175 million)
  - 123 million files (1.9 TB) archived by the Internet Archive from the .PT domain between 1996 and 2007
  - CD ROM with few but interesting sites published in 1996

Oldest Library of Congress site



Choose a topic below, see what's new, or search our Web pages and Gopher menus.

General Information and Publications

Find out about the Library and its mission, special programs and services, information for visitors, publications (including Library Associates and *Civilization Magazine*), employment opportunities, and other general information.

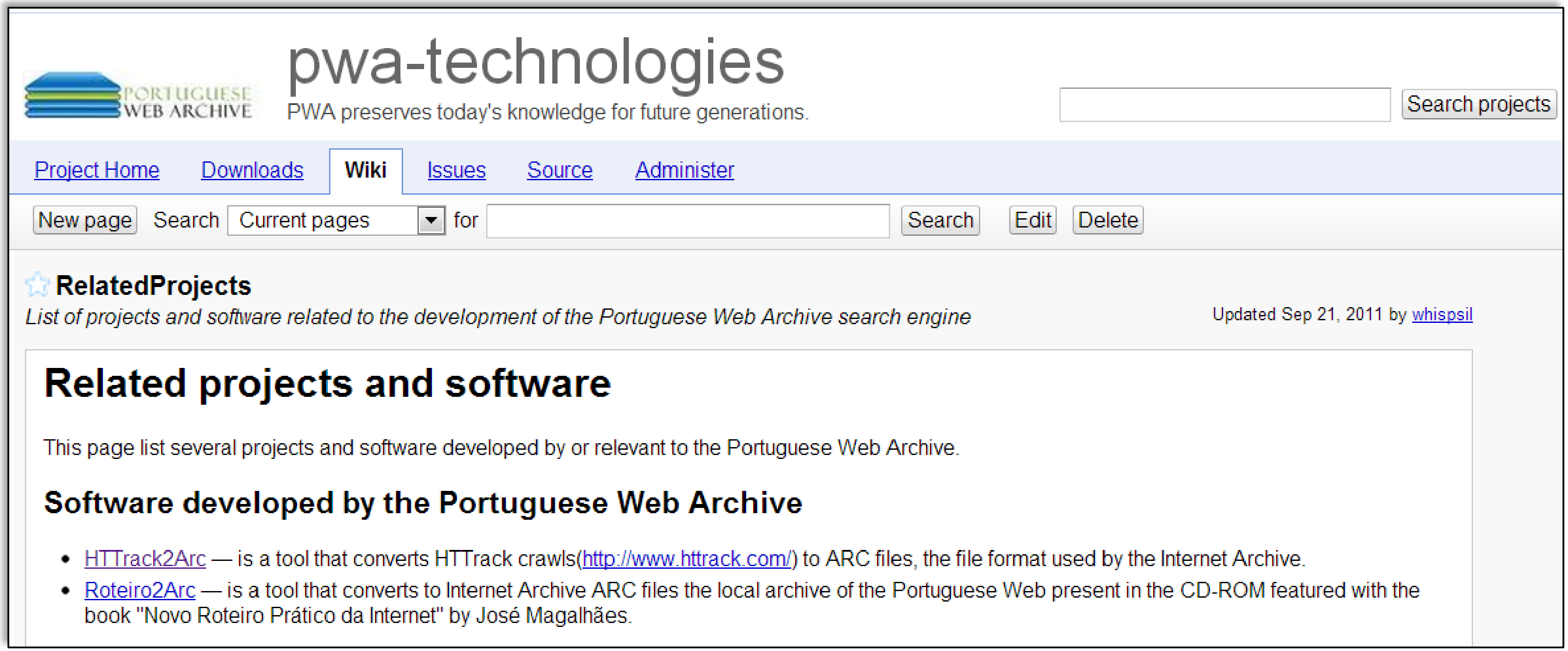
Government, Congress, and Law

Search THOMAS (legislative information), access services of the Law Library of Congress (including the Global Legal Information Network), or locate government information.

Research and Collections Services

Browse historical collections for the National Digital Library (American Memory), visit Library Reading Rooms, access special services for persons with disabilities, and read about Library of Congress cataloging, acquisitions, and preservation operations, policies, and related standards.

Tools to convert saved web files to ARC format



RelatedProjects

List of projects and software related to the development of the Portuguese Web Archive search engine

Updated Sep 21, 2011 by whispsil

Related projects and software

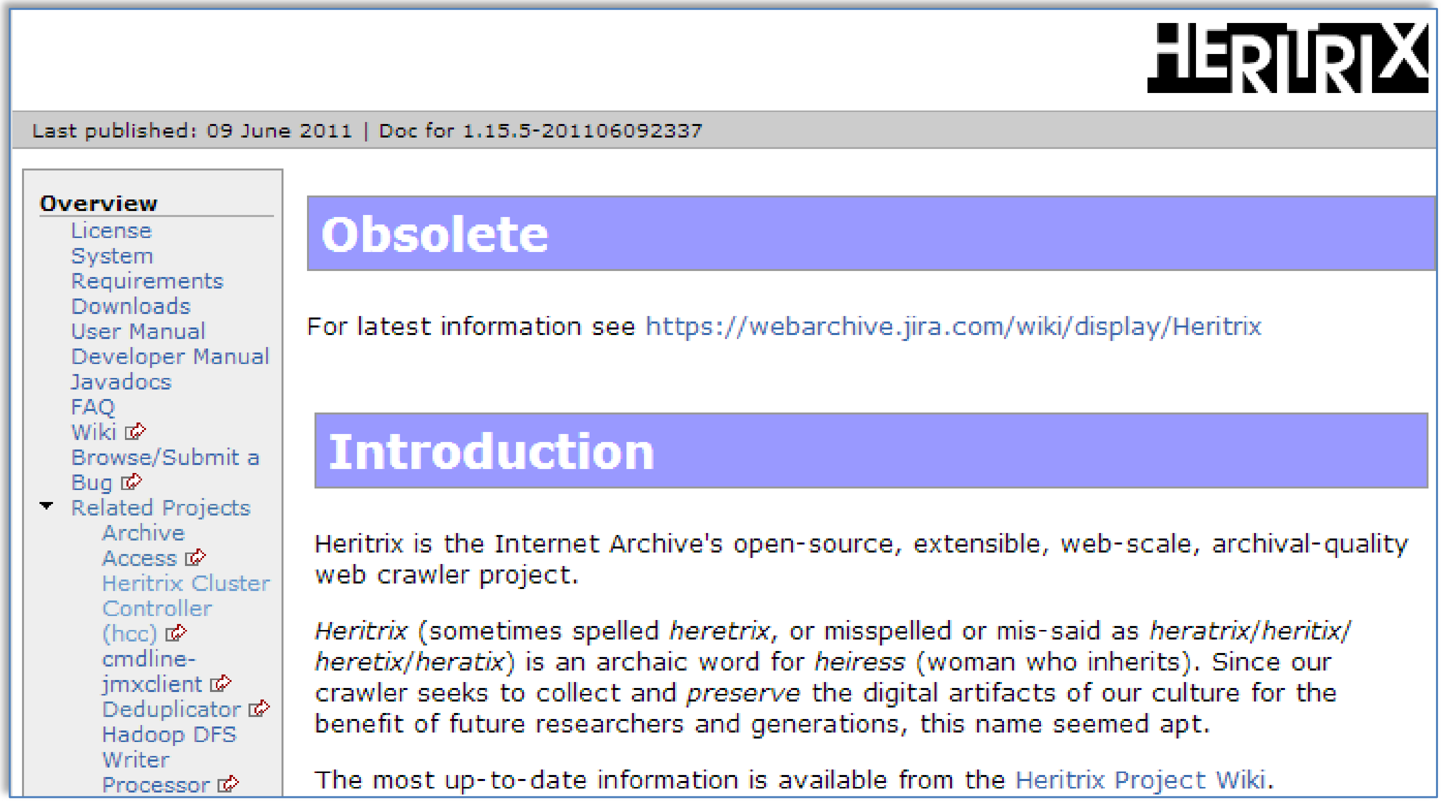
This page list several projects and software developed by or relevant to the Portuguese Web Archive.

Software developed by the Portuguese Web Archive

- HTTrack2Arc — is a tool that converts HTTrack crawls(http://www.httrack.com/) to ARC files, the file format used by the Internet Archive.
- Roteiro2Arc — is a tool that converts to Internet Archive ARC files the local archive of the Portuguese Web present in the CD-ROM featured with the book "Novo Roteiro Prático da Internet" by José Magalhães.

- “Dead” archived collections became searchable and accessible

Crawling the live-web since 2007



Last published: 09 June 2011 | Doc for 1.15.5-201106092337

Overview
- License
- System Requirements
- Downloads
- User Manual
- Developer Manual
- Javadocs
- FAQ
- Wiki
- Browse/Submit a Bug
- Related Projects
  - Archive Access
  - Heritrix Cluster Controller (hcc)
  - cmdline-jmxclient
  - Deduplicator
  - Hadoop DFS Writer Processor

Obsolete

For latest information see https://webarchive.jira.com/wiki/display/Heritrix

Introduction

Heritrix is the Internet Archive's open-source, extensible, web-scale, archival-quality web crawler project.

Heritrix (sometimes spelled *heretrix*, or misspelled or mis-said as *heratrix*/*heritix*/*heretix*/*heratix*) is an archaic word for *heiress* (woman who inherits). Since our crawler seeks to collect and *preserve* the digital artifacts of our culture for the benefit of future researchers and generations, this name seemed apt.

The most up-to-date information is available from the Heritrix Project Wiki.

- Heritrix 1.14.3 configured based on previous experience crawling the Portuguese Web
  - 10 000 URLs per site
  - Maximum file size of 10 MB
  - Courtesy pause of 10 seconds
  - All media types
  - …

Trimestral broad crawls

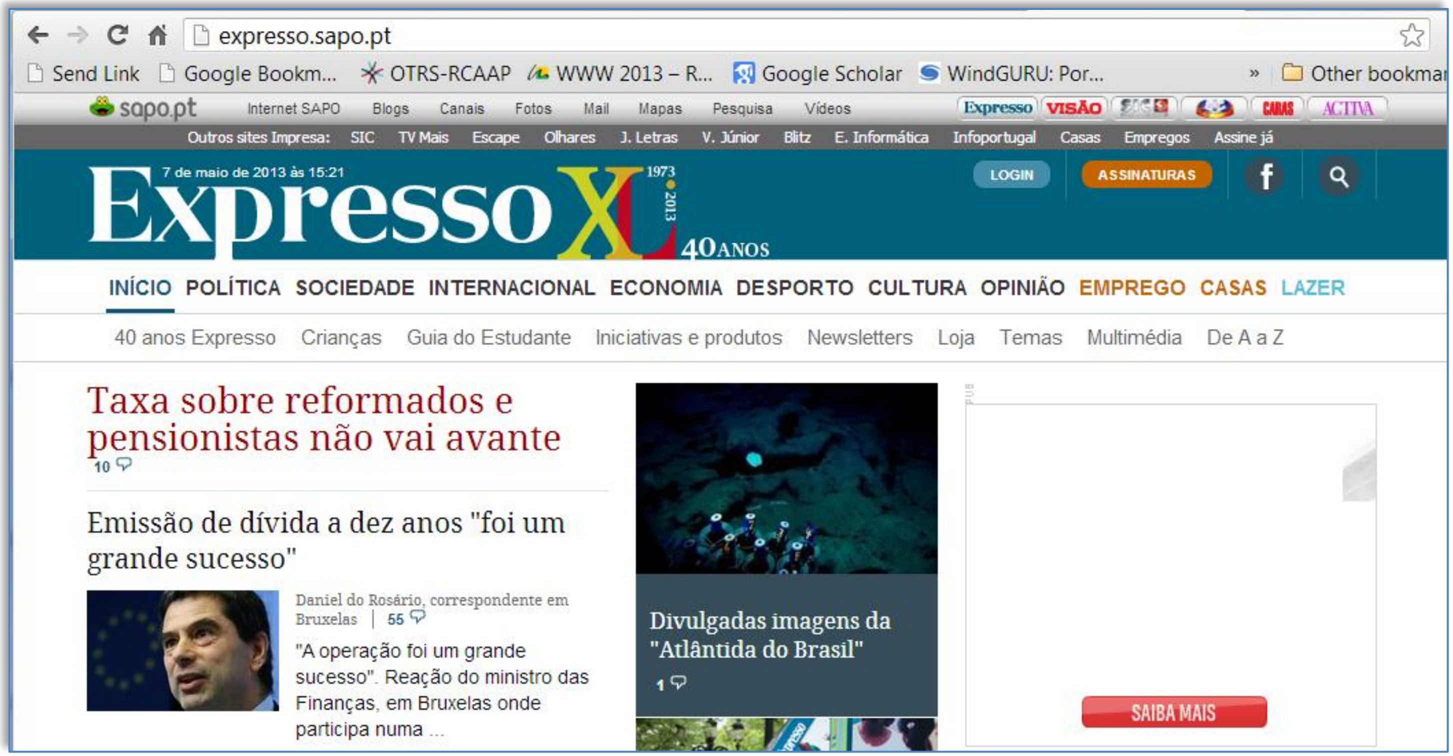
- Includes Portuguese speaking domains (except Brazil)
- 500 000 seeds
  - ccTLD domain listings (.PT, .CV, .AO)
  - User submissions
  - Web directories
  - Home pages of previous crawl
- 78 million files per crawl (5.9 TB)
- New sites from allowed domains are crawled

Daily selective crawls

- 359 online publications selected with the National Library of Portugal
  - Online news and magazines
- Begins at 16:00 to avoid site overload
- Reaches 90% at 7:00
- 764 000 files per day (42 GB)

Problems with daily crawls

The URLs of the publications change frequently



Taxa sobre reformados e pensionistas não vai avante

10

Emissão de dívida a dez anos "foi um grande sucesso"

Daniel do Rosário, correspondente em Bruxelas | 55

"A operação foi um grande sucesso". Reação do ministro das Finanças, em Bruxelas onde participa numa ...

Divulgadas imagens da "Atlântida do Brasil"

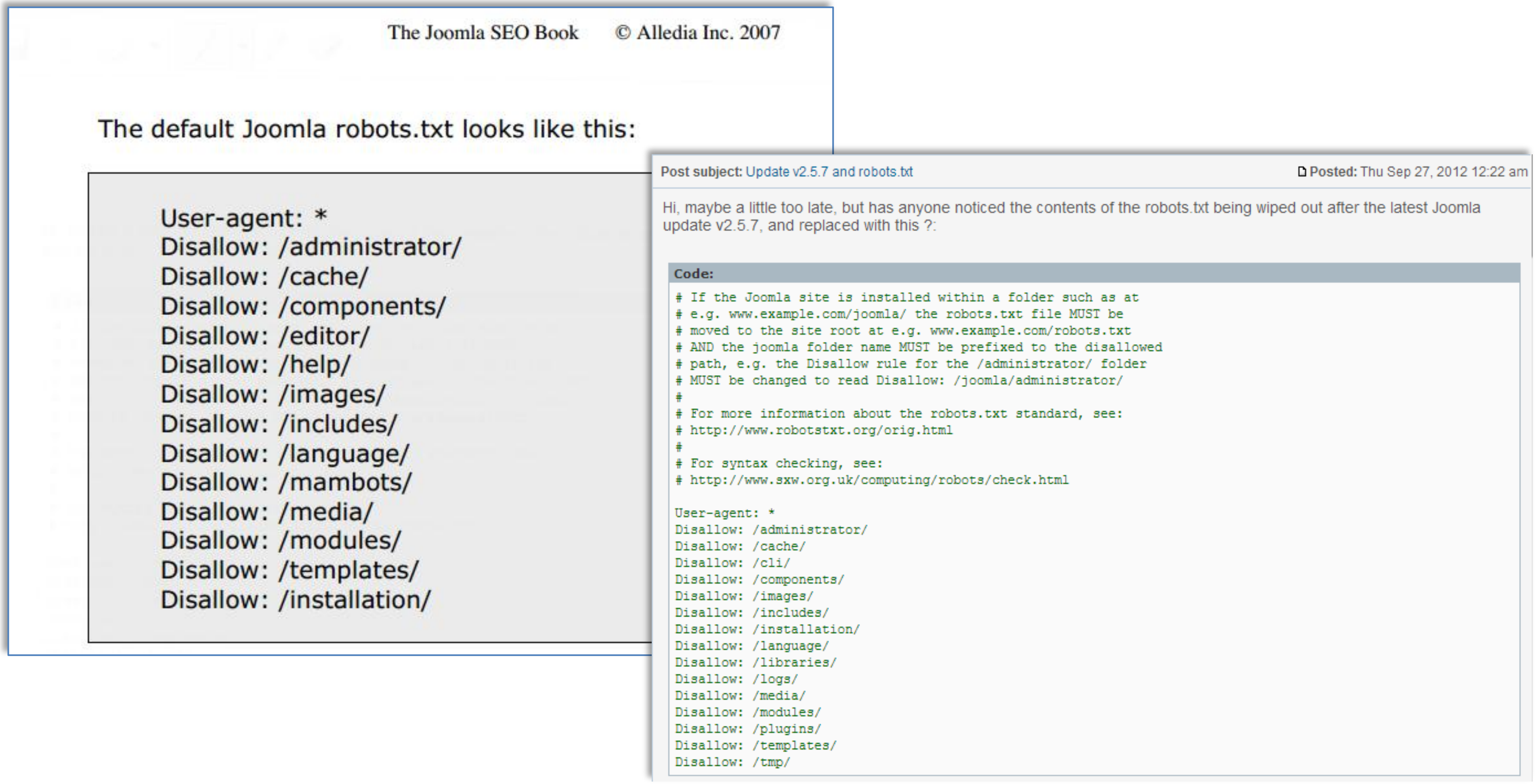
1

- Expresso newspaper since 2008
  - www.expresso.pt, aeiou.expresso.pt, expresso.clix.pt, online.expresso.pt, expresso.sapo.pt
  - Crawl all domains: many duplicates
  - Crawl only new domain: miss legacy content on previous domains
- Must be periodically validated by humans

Default Robots.txt of Content Management Systems forbid crawling images

- Developers of popular Content Management Systems are not aware of web archiving
  - Mambo, Joomla
- Search engines only need the textual content

Joomla robots.txt forbids crawling images since 2007



The Joomla SEO Book © Alledia Inc. 2007

The default Joomla robots.txt looks like this:

```
User-agent: *
Disallow: /administrator/
Disallow: /cache/
Disallow: /components/
Disallow: /editor/
Disallow: /help/
Disallow: /images/
Disallow: /includes/
Disallow: /language/
Disallow: /mambots/
Disallow: /media/
Disallow: /modules/
Disallow: /templates/
Disallow: /installation/
```

Post subject: Update v2.5.7 and robots.txt **Posted:** Thu Sep 27, 2012 12:22 am

Hi, maybe a little too late, but has anyone noticed the contents of the robots.txt being wiped out after the latest Joomla update v2.5.7, and replaced with this ?:

Code:

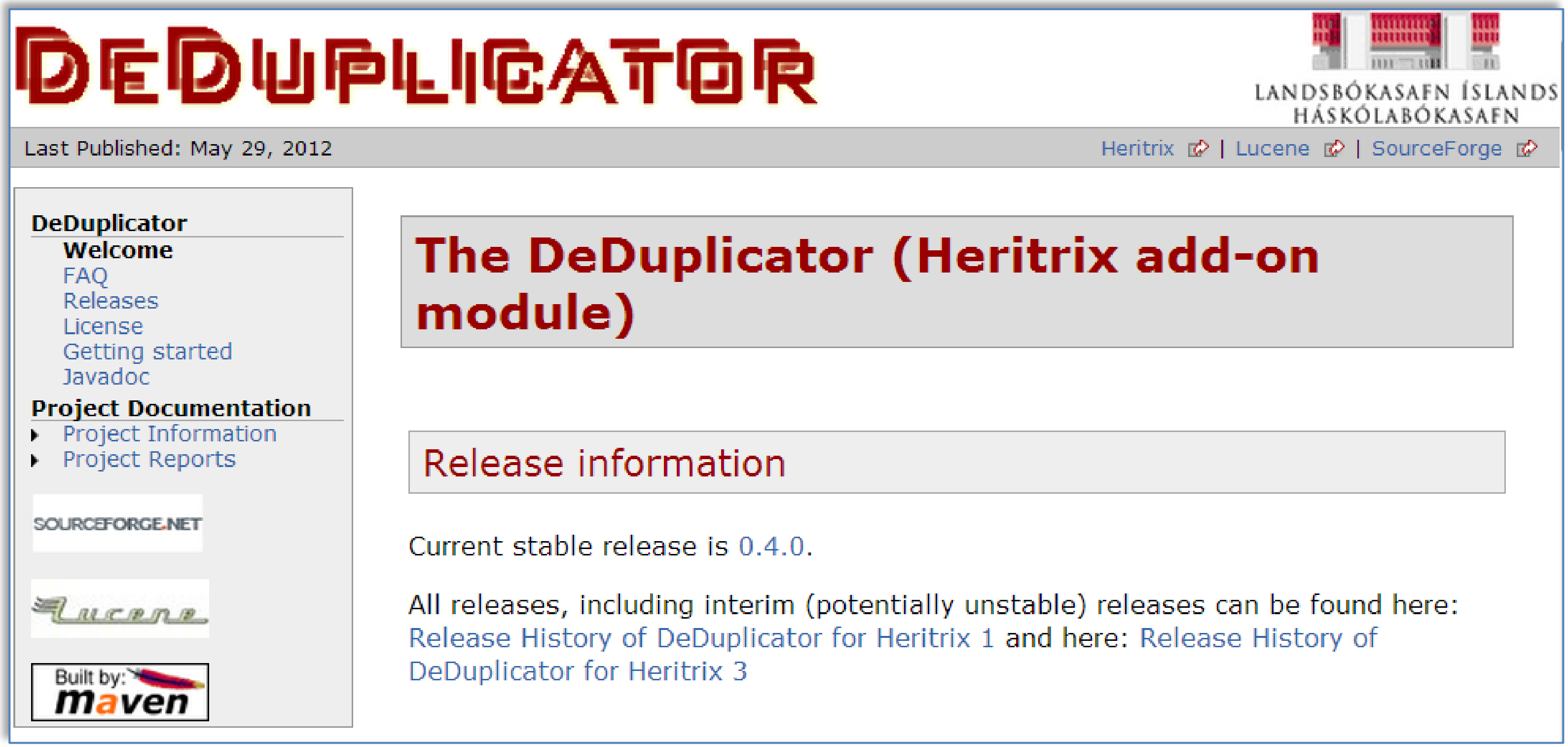
```
# If the Joomla site is installed within a folder such as at
# e.g. www.example.com/joomla/ the robots.txt file MUST be
# moved to the site root at e.g. www.example.com/robots.txt
# AND the joomla folder name MUST be prefixed to the disallowed
# path, e.g. the Disallow rule for the /administrator/ folder
# MUST be changed to read Disallow: /joomla/administrator/
#
# For more information about the robots.txt standard, see:
# http://www.robotstxt.org/orig.html
#
# For syntax checking, see:
# http://www.sxw.org.uk/computing/robots/check.html

User-agent: *
Disallow: /administrator/
Disallow: /cache/
Disallow: /cli/
Disallow: /components/
Disallow: /images/
Disallow: /includes/
Disallow: /installation/
Disallow: /language/
Disallow: /libraries/
Disallow: /logs/
Disallow: /media/
Disallow: /modules/
Disallow: /plugins/
Disallow: /templates/
Disallow: /tmp/
```

- Joomla has been widely used

Attempt to raise awareness

- Contacted webmasters of the selected publications by email
  - Only 10% returned feedback
- Some of them did not know they had robots exclusion rules on their sites



Last Published: May 29, 2012

The DeDuplicator (Heritrix add-on module)

Release information

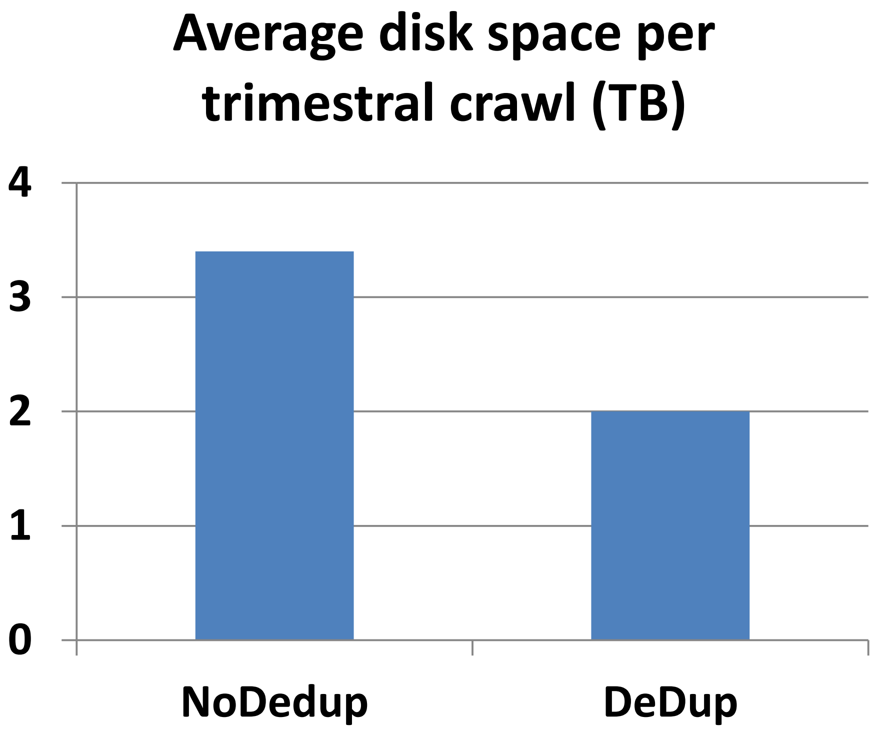
Current stable release is 0.4.0.

All releases, including interim (potentially unstable) releases can be found here: Release History of DeDuplicator for Heritrix 1 and here: Release History of DeDuplicator for Heritrix 3

- Downloads content, computes checksum and compares it with version from the previous crawl
  - Unchanged->Discarded
  - Changed->Stored
- No impact on download rate

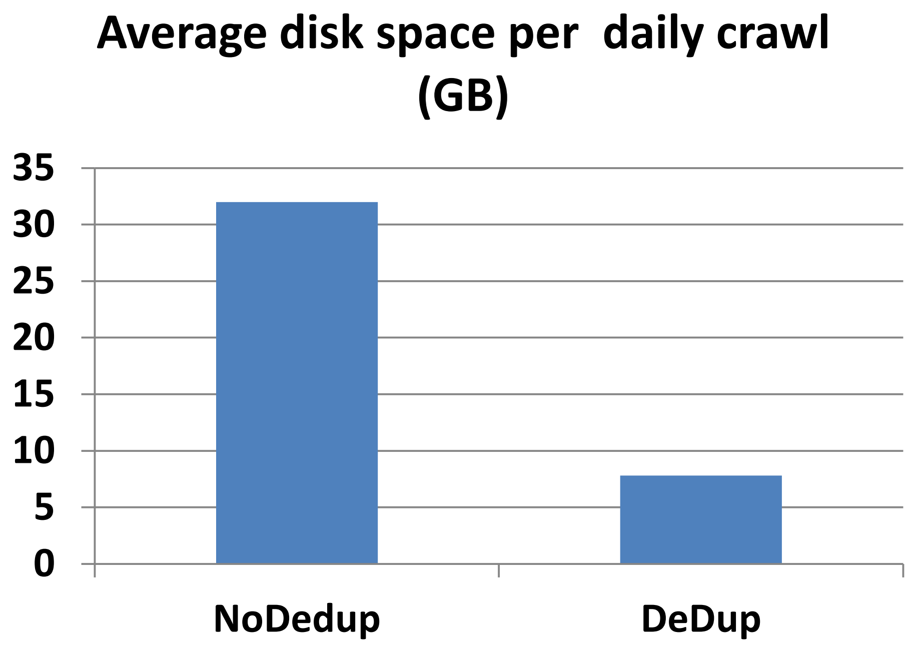
How much space did we save?

Savings on Trimestral crawls



- 41% less disk space to store content
- 1.4 TB saved every 3 months

Savings on Daily crawls



- 76% less disk space to store content
- 24.2 GB saved everyday (8.9 TB/year)

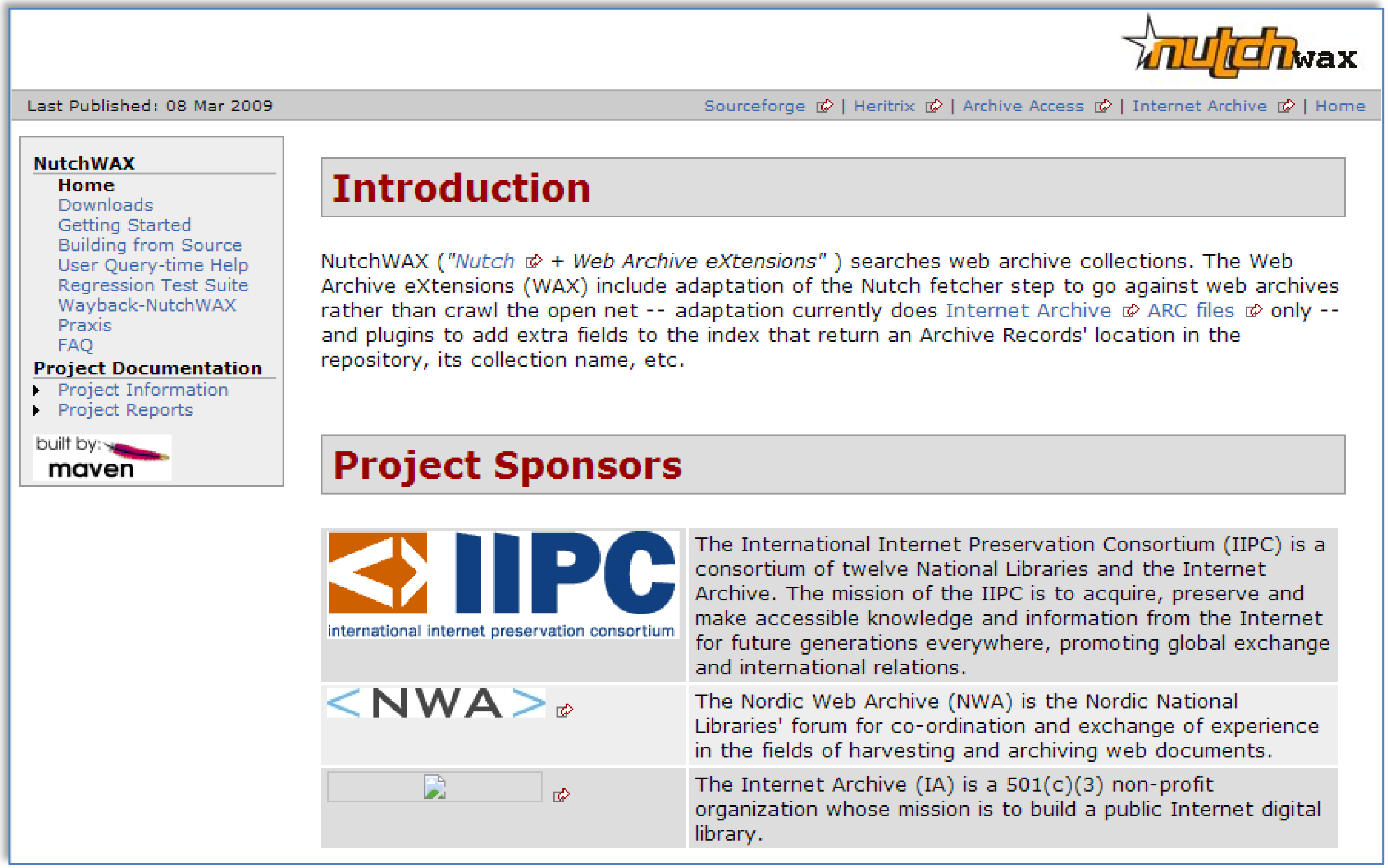
Total savings from using DeDuplicator

26.5 TB/year

Ranking the past Web

Efforts to evaluate and improve search ranking results

NutchWAX as baseline for full-text search



Last Published: 08 Mar 2009

Sourceforge | Heritrix | Archive Access | Internet Archive | Home

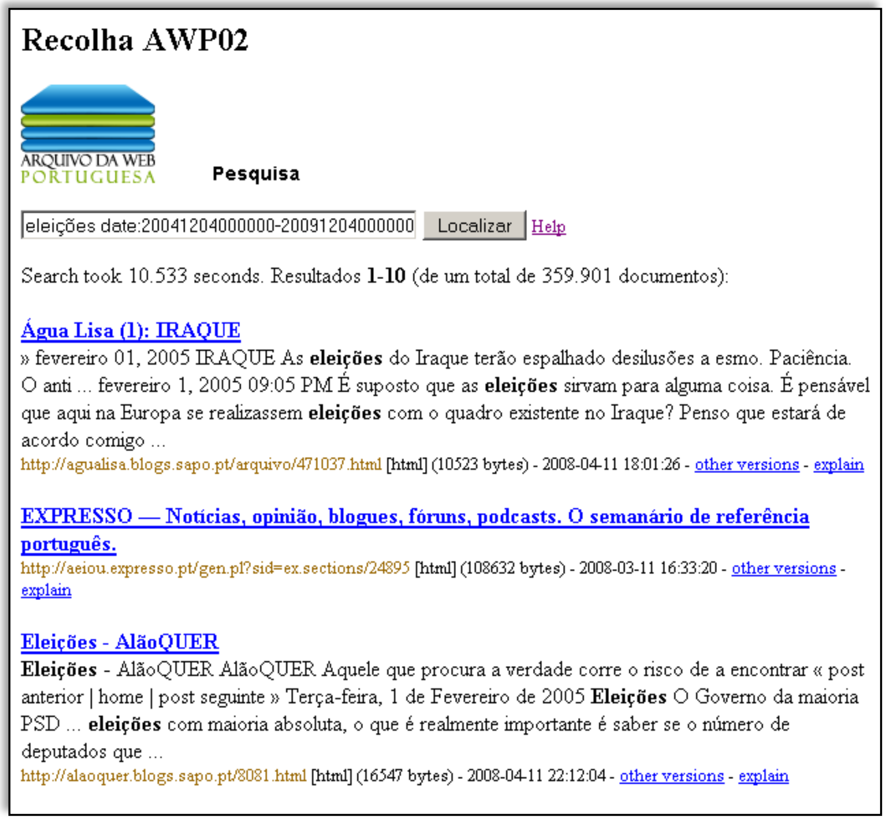
Introduction

NutchWAX (*"Nutch + Web Archive eXtensions"*) searches web archive collections. The Web Archive eXtensions (WAX) include adaptation of the Nutch fetcher step to go against web archives rather than crawl the open net -- adaptation currently does Internet Archive ARC files only -- and plugins to add extra fields to the index that return an Archive Records' location in the repository, its collection name, etc.

Project Sponsors

IIPC international internet preservation consortium	The International Internet Preservation Consortium (IIPC) is a consortium of twelve National Libraries and the Internet Archive. The mission of the IIPC is to acquire, preserve and make accessible knowledge and information from the Internet for future generations everywhere, promoting global exchange and international relations.
< NWA >	The Nordic Web Archive (NWA) is the Nordic National Libraries' forum for co-ordination and exchange of experience in the fields of harvesting and archiving web documents.
	The Internet Archive (IA) is a 501(c)(3) non-profit organization whose mission is to build a public Internet digital library.

Users were not satisfied with NutchWAX search



Recolha AWP02

eleições date:20041204000000-20091204000000

Search took 10.533 seconds. Resultados **1-10** (de um total de 359.901 documentos):

Água Lisa (1): IRAQUE
» fevereiro 01, 2005 IRAQUE As **eleições** do Iraque terão espalhado desilusões a esmo. Paciência. O anti ... fevereiro 1, 2005 09:05 PM É suposto que as **eleições** sirvam para alguma coisa. É pensável que aqui na Europa se realizassem **eleições** com o quadro existente no Iraque? Penso que estará de acordo comigo ...
http://agualisa.blogs.sapo.pt/arquivo/471037.html [html] (10523 bytes) - 2008-04-11 18:01:26 - other versions - explain

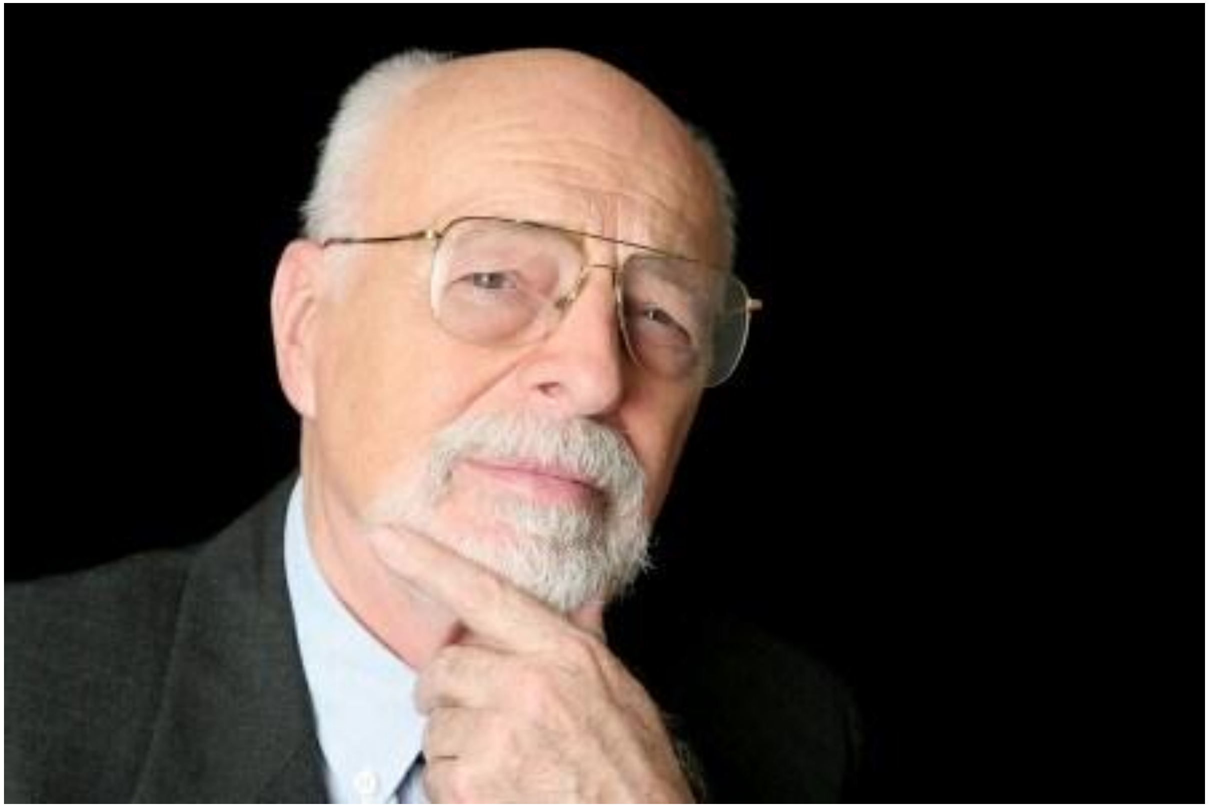
EXPRESSO — Notícias, opinião, blogues, fóruns, podcasts. O semanário de referência português.
http://aeiou.expresso.pt/gen.pl?sid=ex.sections/24895 [html] (108632 bytes) - 2008-03-11 16:33:20 - other versions - explain

Eleições - AlãoQUER
Eleições - AlãoQUER AlãoQUER Aquele que procura a verdade corre o risco de a encontrar « post anterior | home | post seguinte » Terça-feira, 1 de Fevereiro de 2005 **Eleições** O Governo da maioria PSD ... **eleições** com maioria absoluta, o que é realmente importante é saber se o número de deputados que ...
http://alaoquer.blogs.sapo.pt/8081.html [html] (16547 bytes) - 2008-04-11 22:12:04 - other versions - explain

- Unpolished interface
- Slow results
  - 40M URLs, >20s
- Low relevance for search results

Developed a new web archive search system

- Quicker response times
- Improve relevance for search results



“Improved relevance”?!
How did you evaluate your results?

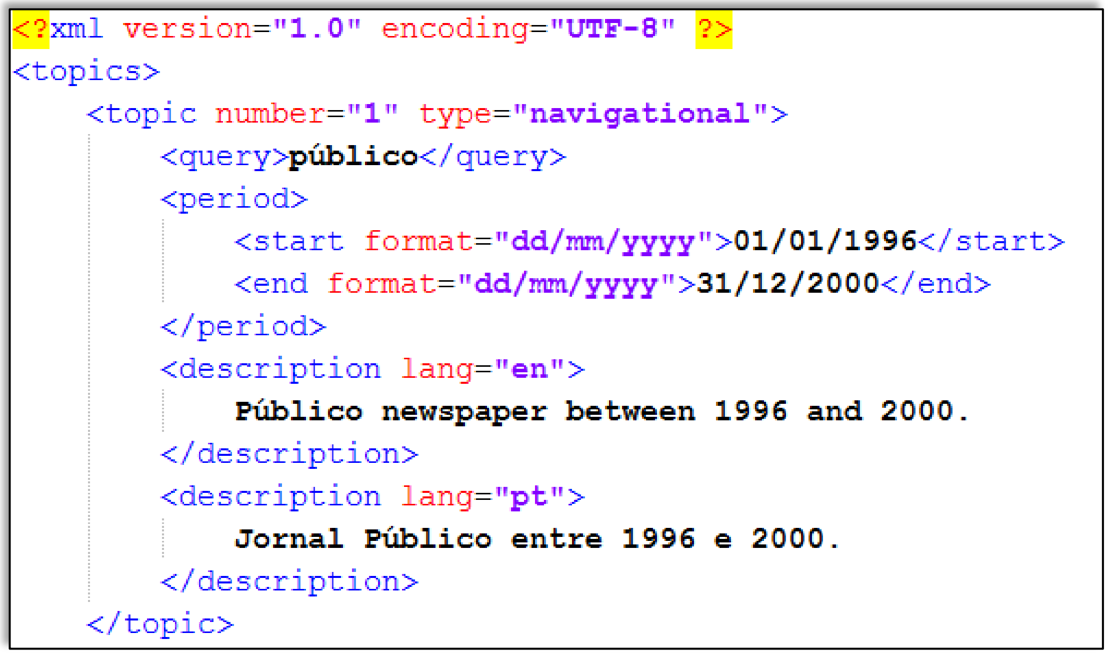
Evaluated our web archive search with TREC benchmark

- TD2003, TD 2004 created to evaluate live-web ranking models
- Our initial ranking model
  - Document fields
    - URL, title, body text, anchor text, incoming links
    - **No temporal fields: crawl date**
  - Ranking features
    - Lucene (based on TFxIdF), Term distance between query terms and title, content, anchor text
    - **No temporal ranking features: age of the page**
- TREC results were acceptable but relevance of our results was obviously weak
  - Inadequate testing

We built a Web Archive Information Retrieval Test Collection: PWA9609

- Corpus of documents from 1996 to 2009
  - 255 million web pages (8.9 TB)
  - 6 collections: Internet Archive, PWA broad crawls, integrated collections

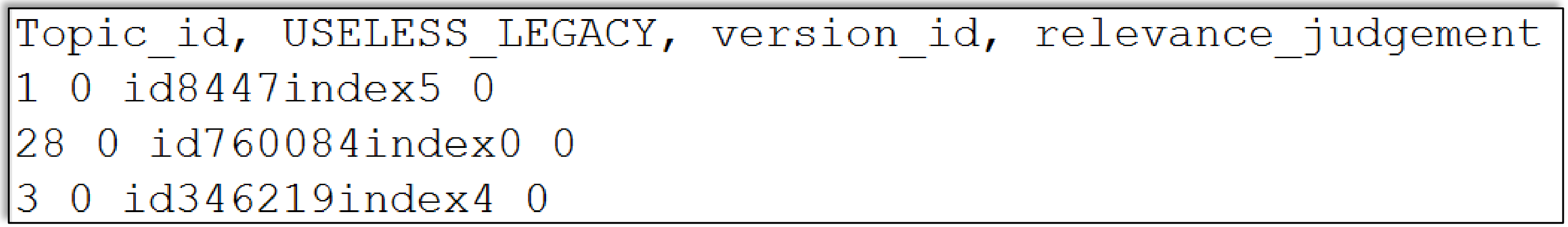
Topics describing users' information needs (topics.xml)



```
<?xml version="1.0" encoding="UTF-8" ?>
<topics>
    <topic number="1" type="navigational">
        <query>público</query>
        <period>
            <start format="dd/mm/yyyy">01/01/1996</start>
            <end format="dd/mm/yyyy">31/12/2000</end>
        </period>
        <description lang="en">
            Público newspaper between 1996 and 2000.
        </description>
        <description lang="pt">
            Jornal Público entre 1996 e 2000.
        </description>
    </topic>
```

- Only navigational topics
  - I need the page of Público newspaper between 1996 and 2000.

Relevance judgments for each topic (qrels)

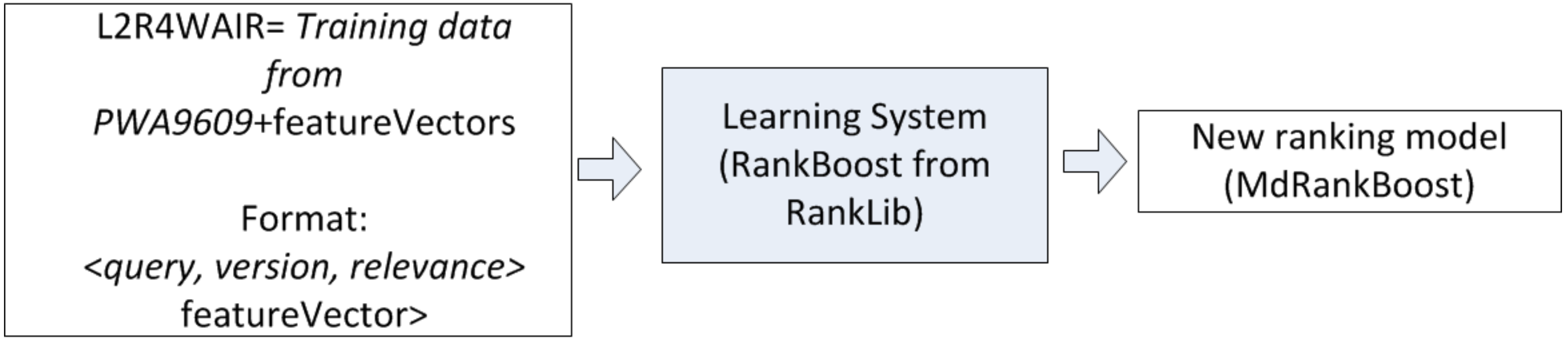


```
Topic_id, USELESS_LEGACY, version_id, relevance_judgement
1 0 id8447index5 0
28 0 id760084index0 0
3 0 id346219index4 0
```

- TREC format to enable reuse of tools

Time-aware ranking models evaluated with the PWA9609 test collection

Time-aware ranking models derived from Learning2Rank



L2R4WAIR= *Training data from PWA9609*+featureVectors

Format:
<*query, version, relevance*> featureVector>

Learning System (RankBoost from RankLib)

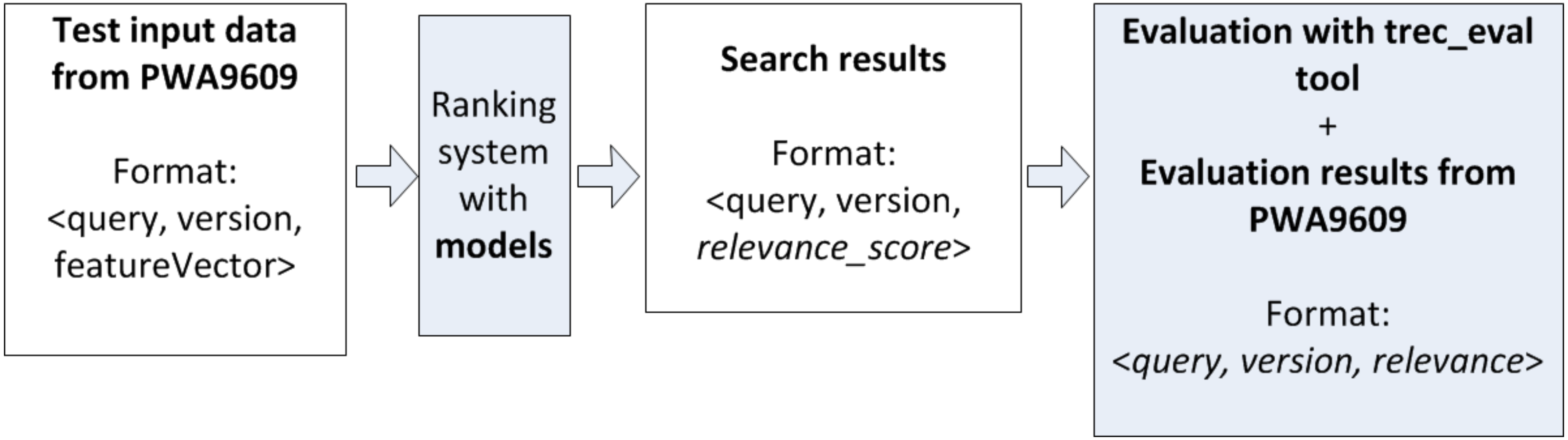
New ranking model (MdRankBoost)

- **MdRankBoost**: RankBoost machine learning algorithm over L2R4WAIR

Time-aware ranking models based on intuition

- **Assumption**: persistent URLs tend to reference persistent content (Gomes, 2006)
- **Intuition**: URLs that persist longer are more relevant
- **TVersions**: higher relevance to URLs with larger number of versions
- **TSpan**: higher relevance to documents with larger time span between first and last version

Evaluation methodology



Test input data from PWA9609

Format:
<query, version, featureVector>

Ranking system with **models**

Search results

Format:
<query, version, *relevance_score*>

Evaluation with trec_eval tool
+
Evaluation results from PWA9609

Format:
<query, version, relevance>

Results

Metric	Time-unaware ranking models	Time-aware ranking models (our proposals)		
	NutchWAX	TVersions	TSpan	**MdRankBoost (L2R)**
nDCG@1	0.250	0.430	0.450	**0.550**
nDCG@10	0.174	0.202	0.193	**0.555**
Precision@1	0.320	0.500	0.520	**0.600**
Precision@10	0.168	0.172	0.158	**0.194**

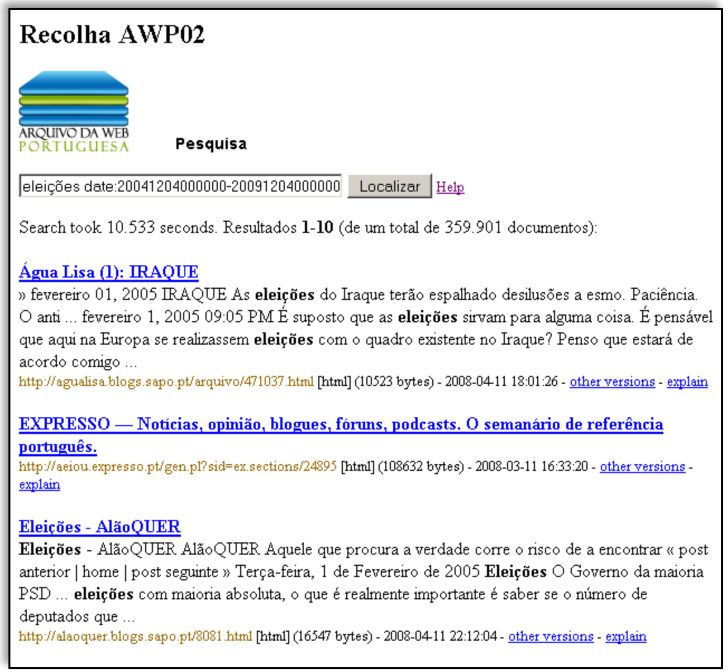
- Temporal L2R approach provided the best results (MdRankBoost)
  - 68 features including temporal features
- TVersions and TSpan yield similar results
  - Persistence of URLs influences relevance
- More details: Miguel Costa, Mário J. Silva, Evaluating Web Archive Search Systems, WISE’2012

Future Work

- Temporal L2R (MdRankBoost) provided the most relevant results
  - 68 features take too much effort to compute
  - Need feature selection
- Extend test collection to include informational queries and re-evaluate ranking models
  - Who won the 2001 Portuguese elections?

Designing user interface

NutchWAX (2007) vs. PWA (2012)



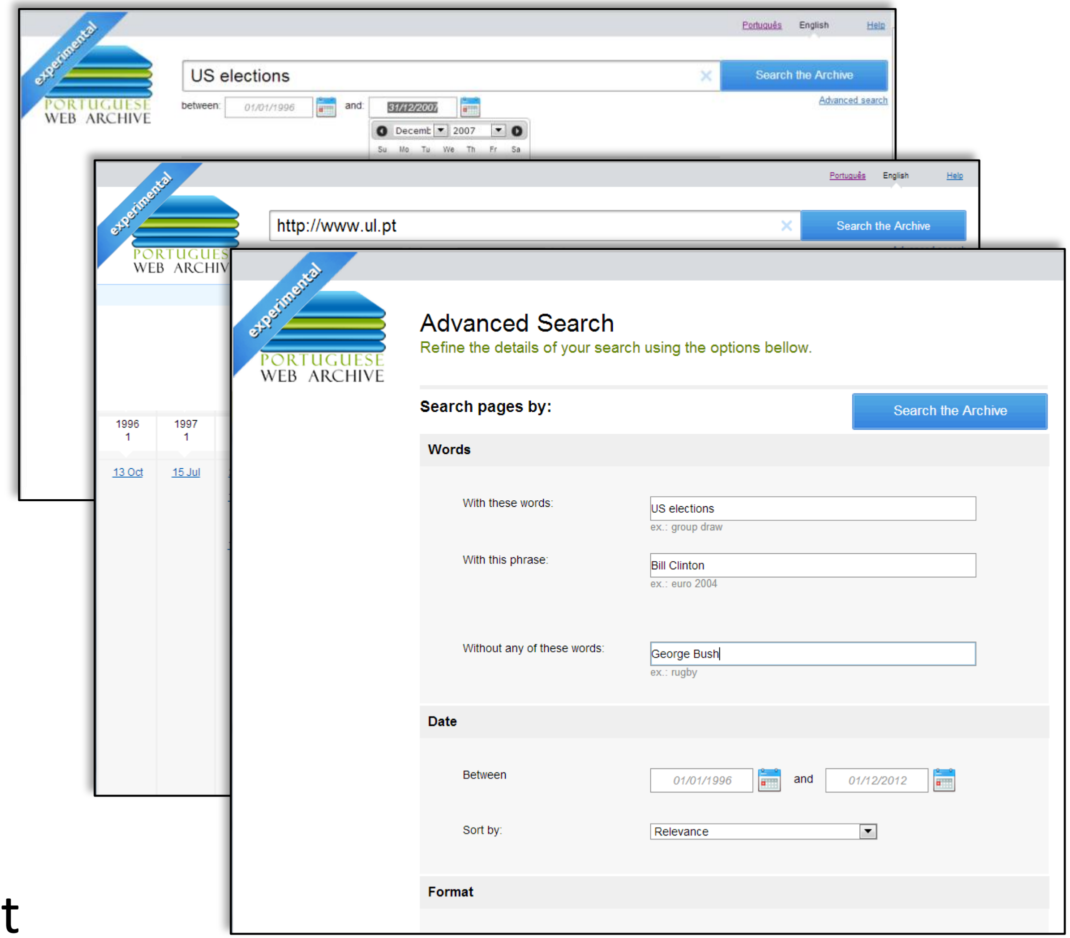
eleições date:20041204000000-20091204000000

Search took 10.533 seconds. Resultados **1-10** (de um total de 359.901 documentos):

[Água Lisa (1): IRAQUE](http://agualisa.blogs.sapo.pt/arquivo/471037.html)
» fevereiro 01, 2005 IRAQUE As **eleições** do Iraque terão espalhado desilusões a esmo. Paciência. O anti ... fevereiro 1, 2005 09:05 PM É suposto que as **eleições** sirvam para alguma coisa. É pensável que aqui na Europa se realizassem **eleições** com o quadro existente no Iraque? Penso que estará de acordo comigo ...
http://agualisa.blogs.sapo.pt/arquivo/471037.html [html] (10523 bytes) - 2008-04-11 18:01:26 - other versions - explain

[EXPRESSO — Notícias, opinião, blogues, fóruns, podcasts. O semanário de referência português.](http://aeiou.expresso.pt/gen.pl?sid=ex.sections/24895)
http://aeiou.expresso.pt/gen.pl?sid=ex.sections/24895 [html] (108632 bytes) - 2008-03-11 16:33:20 - other versions - explain

[Eleições - AlãoQUER](http://alaoquer.blogs.sapo.pt/8081.html)
Eleições - AlãoQUER AlãoQUER Aquele que procura a verdade corre o risco de a encontrar « post anterior | home | post seguinte » Terça-feira, 1 de Fevereiro de 2005 **Eleições** O Governo da maioria PSD ... **eleições** com maioria absoluta, o que é realmente importante é saber se o número de deputados que ...
http://alaoquer.blogs.sapo.pt/8081.html [html] (16547 bytes) - 2008-04-11 22:12:04 - other versions - explain



US elections

between: 01/01/1996 and: 31/12/2007

Decemb 2007

http://www.ul.pt

Advanced Search

Refine the details of your search using the options bellow.

Search pages by:

Words

With these words: US elections
ex.: group draw

With this phrase: Bill Clinton
ex.: euro 2004

Without any of these words: George Bush
ex.: rugby

Date

Between 01/01/1996 and 01/12/2012

Sort by: Relevance

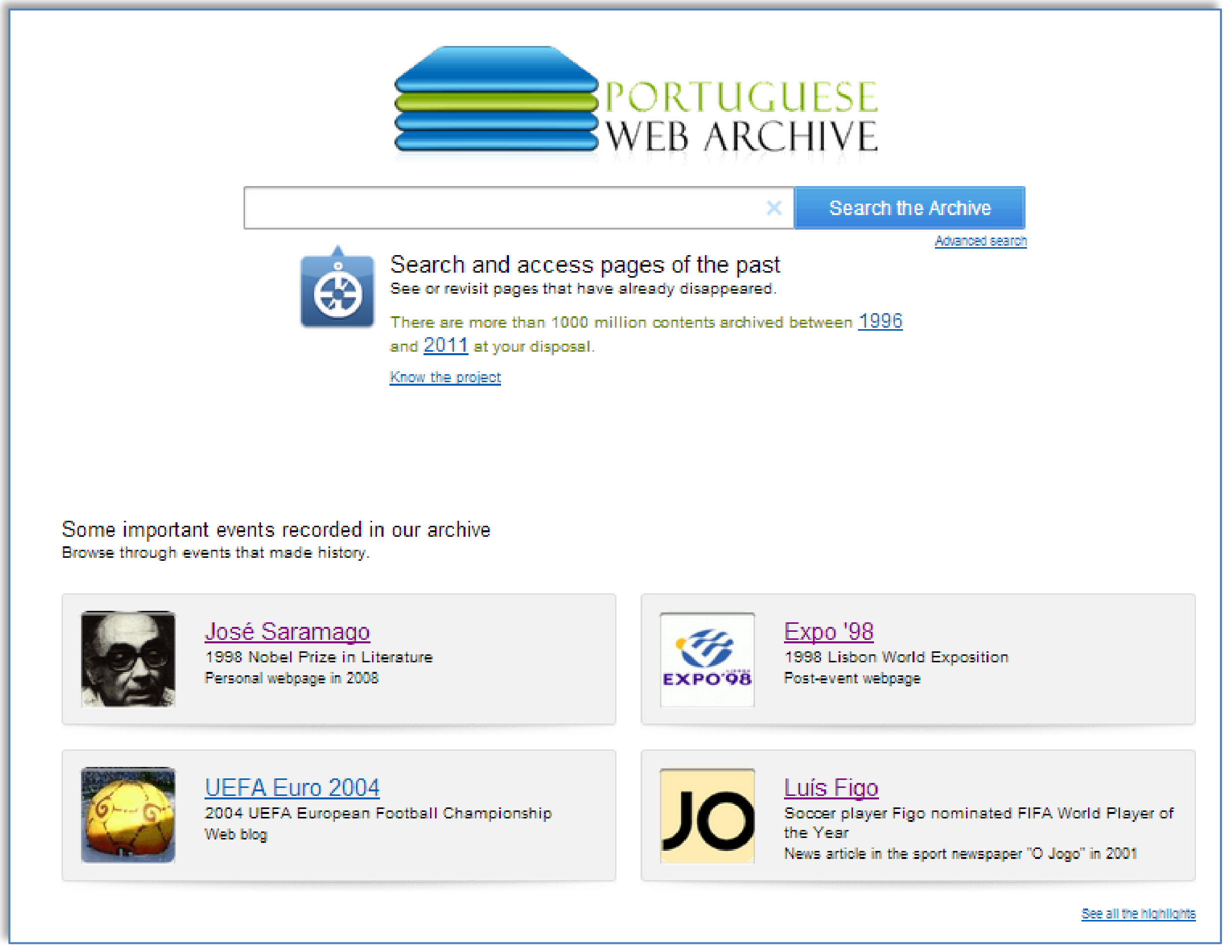
Format

- Internationalization support
- New graphical design
- Advanced search user interface
- 71% overall user satisfaction from rounds of usability testing

Observations from usability testing

USABILITY TESTING
No! That's not how you're supposed to use it!

Searching the past web is a confusing concept



Search and access pages of the past

See or revisit pages that have already disappeared.

There are more than 1000 million contents archived between 1996 and 2011 at your disposal.

Know the project

Some important events recorded in our archive

Browse through events that made history.

José Saramago
1998 Nobel Prize in Literature
Personal webpage in 2008

Expo '98
1998 Lisbon World Exposition
Post-event webpage

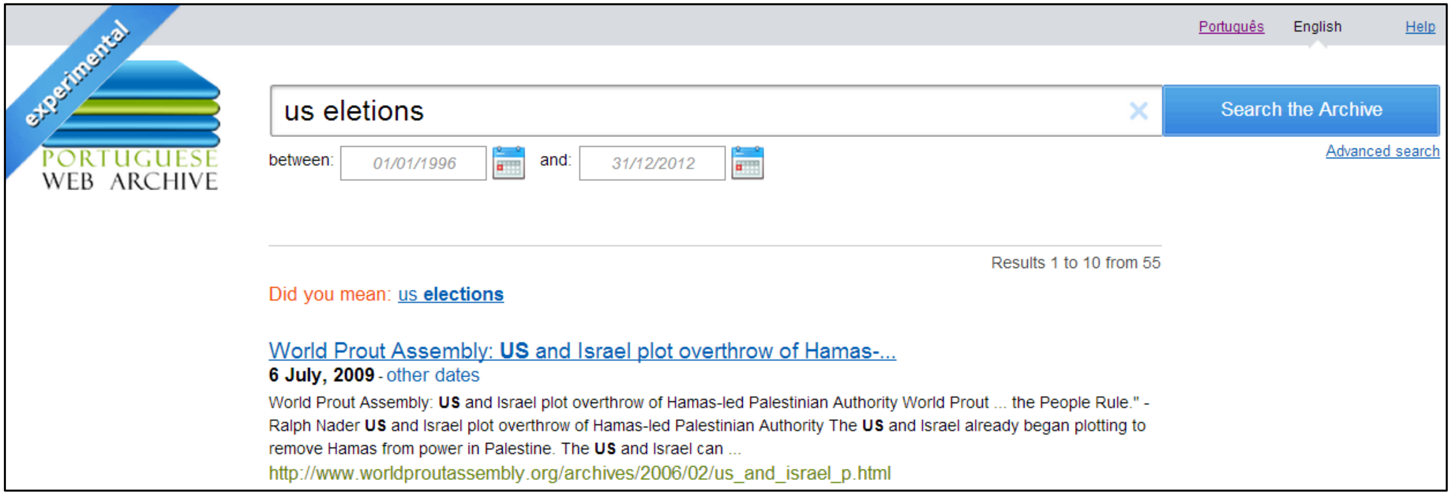
UEFA Euro 2004
2004 UEFA European Football Championship
Web blog

Luís Figo
Soccer player Figo nominated FIFA World Player of the Year
News article in the sport newspaper "O Jogo" in 2001

See all the highlights

- Understanding web archiving requires being techie
- Provide examples of web-archived pages

Users are addicted to query suggestions



us eletions

between: 01/01/1996 and: 31/12/2012

Results 1 to 10 from 55

Did you mean: us **elections**

World Prout Assembly: **US** and Israel plot overthrow of Hamas-...

6 July, 2009 - other dates

World Prout Assembly: **US** and Israel plot overthrow of Hamas-led Palestinian Authority World Prout ... the People Rule." - Ralph Nader **US** and Israel plot overthrow of Hamas-led Palestinian Authority The **US** and Israel already began plotting to remove Hamas from power in Palestine. The **US** and Israel can ...

http://www.worldproutassembly.org/archives/2006/02/us_and_israel_p.html

- Developed query suggestions mechanism for web archive search

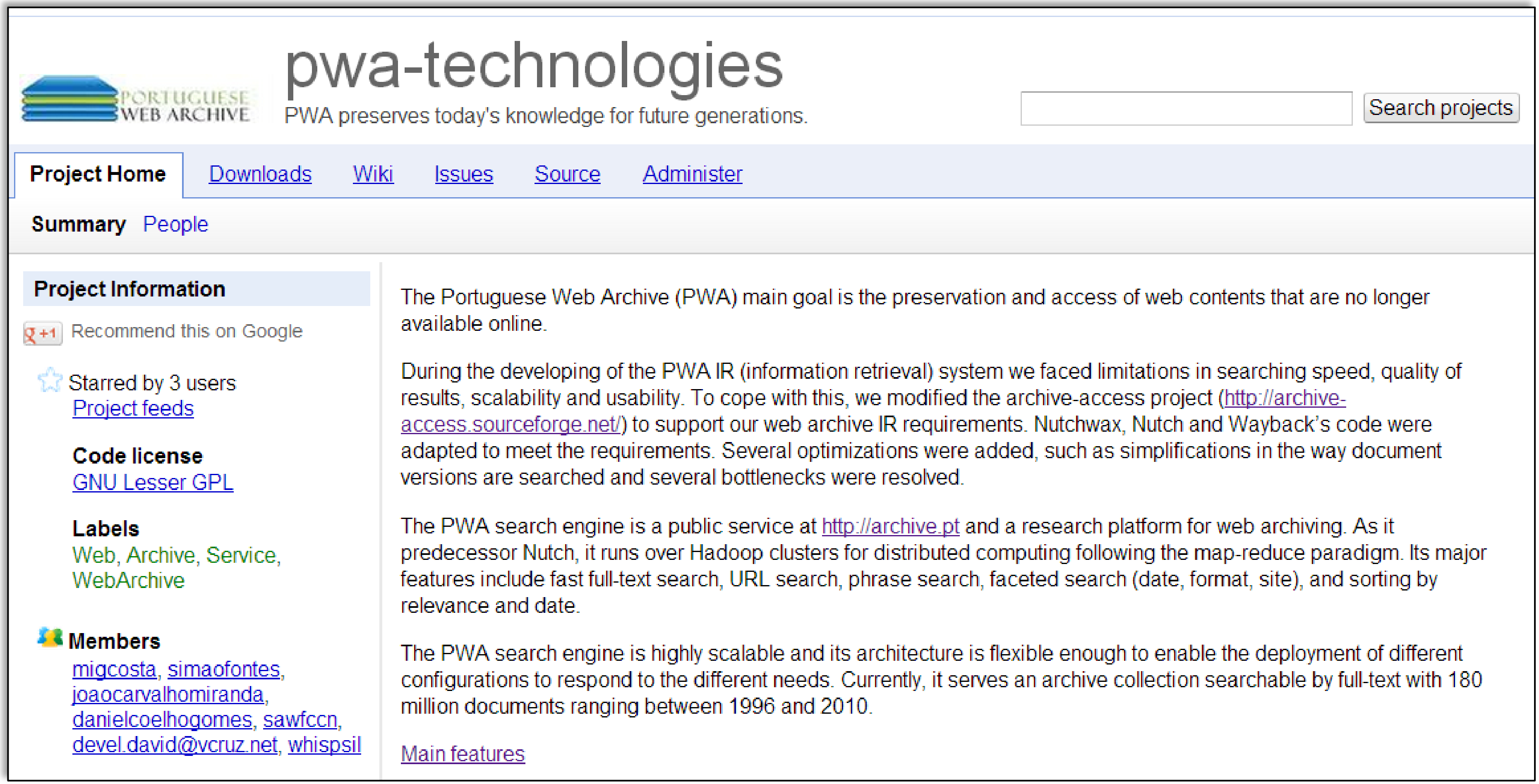
Users “google” the past

- Users search web archives replicating their behavior from live-web search engines
- Users input queries on the first input box that they find
  - Search system must identify query type (URL or full-text) and present corresponding results
- Provide additional tutorials and contextual help

Conclusions

- Must raise awareness about web archiving among users and developers
- Time aware ranking models are crucial to search web archives
- We would like to collaborate with other organizations
    - Project proposals online

All our source code and test collections are freely available



pwa-technologies

PWA preserves today's knowledge for future generations.

Project Information

Recommend this on Google

Starred by 3 users

Project feeds

Code license

GNU Lesser GPL

Labels

Web, Archive, Service, WebArchive

Members

migcosta, simaofontes, joaocarvalhomiranda, danielcoelhogomes, sawfccn, devel.david@vcruz.net, whispsil

The Portuguese Web Archive (PWA) main goal is the preservation and access of web contents that are no longer available online.

During the developing of the PWA IR (information retrieval) system we faced limitations in searching speed, quality of results, scalability and usability. To cope with this, we modified the archive-access project (http://archive-access.sourceforge.net/) to support our web archive IR requirements. Nutchwax, Nutch and Wayback's code were adapted to meet the requirements. Several optimizations were added, such as simplifications in the way document versions are searched and several bottlenecks were resolved.

The PWA search engine is a public service at http://archive.pt and a research platform for web archiving. As it predecessor Nutch, it runs over Hadoop clusters for distributed computing following the map-reduce paradigm. Its major features include fast full-text search, URL search, phrase search, faceted search (date, format, site), and sorting by relevance and date.

The PWA search engine is highly scalable and its architecture is flexible enough to enable the deployment of different configurations to respond to the different needs. Currently, it serves an archive collection searchable by full-text with 180 million documents ranging between 1996 and 2010.

Main features

Visit me at the Demo lobby during the conference

Thanks.

PORTUGUESE
WEB ARCHIVE

www.archive.pt

daniel.gomes@fccn.pt