context
Internet Archive

Wow!
•Crawling the public Web since 1996
•Over 280 billion URLs
•Archive searchable by URL via Wayback Machine, which contains 5 petabytes of crawl data from 1996-2012
•Wayback Machine gets 1,000 visits per second
•Web archives from 1996-2007 backed up at mirror site at Bibliotheca Alexandrina
•Also collects millions of digitized books, movies, audio and concert recordings
•Over 10 Petabytes of data in total

But!
•IA web collections vast but not comprehensive
•General Internet crawls take about 3 months, many websites change faster or are short-lived or not well-linked/discoverable
•Depth of capture of individual websites varies
•Archive too huge to be indexed for full-text search
•Robots.txt restrictions are obeyed, so many sites fully or partially blocked from archiving
A GLOBAL NETWORK OF EXPERTS ARCHIVING THE WEB FOR FUTURE GENERATIONS.

Learn More About IIPC

INTERNATIONAL

WHY ARCHIVE THE WEB?
Information published on the World Wide Web today will be the primary resources for future projects.
44 Members

Member Type
- national/regional libraries 29
- university libraries 7
- other non-profits 4
- commercial 2
- archives 2

Member Region
- Europe 23
- North America 13
  - [Internet Archive, CDL, LC, US GPO, Columbia, UNT, GWU, LANL, ODU, 3 in Canada]
- Asia 4
- Oceania 2
- North Africa 2
Types of web archiving

Domain crawls
  National domains (.pt, .fr, .at)
  National/regional/local government domains (.gov.uk, state.nc.us)
  University domains (columbia.edu)

Thematic/selective collections

Event-based collections

Commercial/legal/compliance

Personal digital archiving
List of Web archiving initiatives

From Wikipedia, the free encyclopedia

This article uses bare URLs for citations. Please consider adding full citations so that the article remains verifiable. Several templates and the Reflinks tool are available to assist in formatting. (Reflinks documentation) (February 2012)

This page contains a list of Web archiving initiatives worldwide. For easier reading, the information is divided in three tables: web archiving initiatives, archived data and access methods.

This Wikipedia page was originally generated from the results obtained for the research paper A survey on web archiving initiatives[1] published by the Portuguese Web Archive[2] team.

<table>
<thead>
<tr>
<th>Web archiving initiatives</th>
<th>[edit]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td><strong>Country</strong></td>
</tr>
<tr>
<td>Full-time</td>
<td>Part-time</td>
</tr>
<tr>
<td>The PANDORA Archive which takes a selective approach is a collaborative program of 11</td>
<td></td>
</tr>
</tbody>
</table>
Welcome to the UK Web Archive

Thousands of UK websites have been collected since 2004 and the Archive is growing fast.

Here you can see how sites have changed over time, locate information no longer available on the live Web and observe the unfolding history of a spectrum of UK activities represented online. Sites that no longer exist elsewhere are found here and those yet to be archived can be saved for the future by nominating them.

The Archive contains sites that reflect the rich diversity of lives and interests throughout the UK. Search is by Title of Website, Full Text or URL, or browse by Subject, Special Collection or Alphabetical List.

Cookies on the UK Web Archive website

We use cookies to ensure that we give you the best experience on our website. If you continue without changing your settings, we'll assume that you are happy to receive all cookies on the UK Web Archive website. Click here for more information.

Explore the Special Collections

Special Collections are groups of websites brought together on a particular theme by librarians, curators and other specialists, often working in collaboration with key organisations in the field. They can be events-based (e.g. The Olympic & Paralympic Games 2012), topical (e.g. The Credit Crunch Collection) or subject-oriented (e.g. The British Countryside Collections).

http://www.webarchive.org.uk/ukwa/
The Library of Congress Web Archives (LCWA) is composed of collections of archived web sites selected by subject specialists to represent web-based information on a designated topic. It is part of a continuing effort by the Library to evaluate, select, collect, catalog, provide access to, and preserve digital materials for future generations of researchers. The early development project for Web archives was called MINERVA.

Web Archives Available:

- Crisis in Darfur, Sudan, Web Archive, 2006
- Indian General Elections 2009 Web Archive
- Indonesian General Elections 2009 Web Archive
- Iraq War 2003 Web Archive
- Law Library Legal Blawgs Web Archive
- Library of Congress Manuscript Division Archive of Organizational Web Sites
- Papal Transition 2005 Web Archive
- September 11, 2001 Web Archive
- Single Sites Web Archive
- United States 107th Congress Web Archive
- United States 108th Congress Web Archive
- United States Election 2000 Web Archive
- United States Election 2002 Web Archive
- United States Election 2004 Web Archive
- United States Election 2006 Web Archive
- United States Election 2008 Web Archive
- Visual Image Web Sites Archive

The Library of Congress  >> More Online Collections
August 5, 2011

http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html
Copyright / Legal deposit

Copyright and legal deposit laws affecting right to harvest websites and provide access to archived websites vary from country to country:

- No legal deposit requirement for websites or explicit library exception to allow harvesting websites (USA)
- Legal deposit of websites but no access (Norway)
- Legal deposit of websites but access is restricted to terminals at national libraries (Austria, France)
- No legal deposit requirement but national domain crawls conducted and full online access provided (Portugal)

UK
From 6 April 2013, under the Legal Deposit Libraries (Non-Print Works) Regulations 2013 (UK), the British Library and the other five specified legal deposit libraries are entitled to harvest content from UK-published material on UK websites. Plan is to make results of domain crawls publicly accessible.

European Union
"Right to be forgotten" proposal. Consumer data protection legislation aimed at commercial giants like Google and Facebook, potential problem for digital libraries, archives
Web Resources Collection Program

Overview

Columbia University Libraries and Information Services (CUL) has expanded the scope of its collection development activities to include curated archival collections of freely available Internet resources. The development of this program, self-funded by CUL as of 2013, was made possible by generous support from the Andrew W. Mellon Foundation.

Columbia University Libraries’ commitment to integrating web archiving into ongoing collection development and preservation best practice is informed by collaboration with other research libraries and the broader web archiving community. In 2012 CUL became a member of the International Internet Preservation Consortium (IIPC) and hosted a summit meeting for practitioners on Web Archiving Policies and Practices in the US. Columbia has also recently received new support from the Andrew W. Mellon Foundation for the explicit goal of fostering web archiving collaboration.

Supporting Grant Projects

The Andrew W. Mellon Foundation has provided CUL funding for three grant projects in the area of web archiving:

Collection Building for Web Resources (2008-2009)

Objective: A joint project with the University of Maryland Libraries to develop and test coherent, holistic models for incorporating web content into research library collections.

Web Resources Collection Program Development (2009-2012)
Columbia University Libraries

Organizations

Description: The Columbia University Libraries (CUL) web resources collection program archives selected websites in thematic areas corresponding to existing CUL collection strengths, websites produced by affiliates of Columbia University, and websites from organizations or individuals whose papers or records are held in CUL's physical archives.

Narrow Your Results

Sites and collections from this organization are listed below. Narrow your results at left, or enter a search query below to find a collection, site, specific URL or to search the text of archived webpages.

Subject

Sort By: Count | (A-Z)

Arts & Humanities (3)
Society & Culture (3)
Universities & Libraries (3)
Architecture (1)
City planning (1)

Creator

Sort By: Count | (A-Z)

Avery Architectural and Fine Arts Library, Columbia University Libraries (1)
Burke Library (Union Theological Seminary) (1)
Columbia University (1)
Columbia University Libraries, Center for Human Rights Documentation and Research (1)

Coverage

Sort By: Count | (A-Z)

New York (N.Y.) (1)

http://www.archive-it.org/organizations/304
CUL Web Archive Collections

- Avery Library Historic Preservation and Urban Planning
  60+ websites, semiannual

- Burke Library New York City Religions
  225 websites, semiannual

- Human Rights
  520 websites, quarterly

- Rare Book and Manuscript Library
  30 websites, semiannual

- University Archives
  most of columbia.edu domain, plus 75+ affiliated sites with "external" URLs

- General
Web Crawling

Benefits
- Best case scenario of fully navigable archived version of site reproducing look and feel of original website on particular date
- Demonstrating evolution of a website over time
- Preserving access to websites no longer available on live web
- Ability to provide researchers three primary study objects: document/file; website; "web sphere"

Challenges
- Imperfect capture/rendering of certain content (javascript, flash, password-protected, query-dependent)
- Crawler traps
- Scoping decisions for duplicative content
- Affect of data/document budget on scoping and crawl frequency decisions
Africa Internally Displaced Persons Voice

(AFRICA IDP VOICE)

Raising awareness and Promoting effective protection of IDPs in Africa

Campaigns
Links
Job/Internship
Membership
IDP Legislation
IDP Policies
Donate Now

Copyright Africa Internally Displaced Persons Voice.
For problems or questions regarding this web contact [Africa IDP Voice]
Last updated: 02/01/07.
Oops! Google Chrome could not connect to www.africaidp.org

Suggestions:
- Go to www.africaide.org
- Go to www.africaaid.org
- Try reloading: www.africaidp.org
- Search on Google:

http://www.africaidp.org/ now unavailable
Le Sénégal ratifie la convention de la CEDEAO sur les Armes Légères et de Petit calibre

Le Sénégal devient ainsi, le cinquième pays à rattraper la Convention de la CEDEAO après le Niger, le Mali, le Burkina Faso et la Sierra Leone.
CUL capture of seed placed in Wayback context
Archive-It partners have the option of supplementing their curated collections with earlier captures of their seed URLs from General Wayback crawl.

**Pro:** Users provided a deeper temporal overview of evolution of each seed URL.

**Con:** Complicates permissions; uneven display of curated and uncurated captures; greatly increases collection data size making local indexing slower, costlier; shows history of a given URL, not necessarily of a specific website.

**Best practice=??**

We've so far chosen not to supplement our own captures.
Problem Areas

- Database-driven sites
- Inconsistent URIs
- Dynamic URIs
- Rich Media
- Incremental loading
- Scripted HTML forms
- Embedded content
- Captchas, logins, etc
- User agent sensitive
- Exclusions
  - By Convention
  - By Design
- Server-side scripts
- HTML5 web sockets
- Mobile publishing
AFAD Statement on the 6th Anniversary of Jonas Burgos’ Disappearance
Asian Federation Against Involuntary Disappearances

- Home
- About AFAD
  - History
  - Constitution
  - By-Laws
  - Orientation
  - Member Organizations
- News
- Gallery
- Resources
  - Books
    - Reclaiming Stolen Lives
  - The Voice
    - August 2012 Issue: Hope Springs Eternal
    - March 2011 Issue: Dawn of Hope
    - March 2010 Issue: On the Road to Ratification
    - August 2009 Issue: Where Are You?
    - December 2008 Issue: Dignity and Justice for All of Us
    - May 2008 Issue
    - August 2007 Issue: Presente!

Capture missing stylesheet due to robots.txt restrictions
Incomplete captures
Examples from a "web sphere" within the Avery Library Historic Preservation and Urban Planning collection:

1. The High Line
http://www.thehighline.org/

2. The Lowline
http://thelowline.org/

3. Governors' Island
http://www.govislandpark.com/
Partially successful capture made public: image galleries not fully functional in archived version
Partially successful capture made public: slideshow not fully functional in archived version

http://thelowline.org/
Remote skylights, designed by James Ramsey of Raad Studio, will deliver sunlight below ground using advanced optical systems. Here's how it works.

- Solar collection dish: reflective parabola collects sunlight
- Tracking mechanism: follows the path of the sun throughout the year
- Helio tube: fiberoptic cable channels sunlight through the street to subterrannean subway stations
- Dome: reflects and distributes channelled sunlight
- Green space: underground sunlight sustains plants, trees and grass

Some key content captured successfully
Flash-heavy live site: http://www.govislandpark.com/
Partially successful capture not yet made public: Flash slideshow missing
Crawl scoping

Test crawls demonstrate that selected sites have "crawler traps" or other design elements that capture large amounts of unwanted URLs, draining our data budget, diverting crawler from desirable content, distorting/cluttering search results.

But, selectively scoping out certain URLs from archived websites may have unanticipated replay effects or impede certain types of future research.

Best practice = ??
We choose to scope actively
URL crawl report indicating Calendar "crawler trap"
Next month, next month, next month ...
Using regular expression in Archive-It to block URLs corresponding to unneeded duplicative content
<table>
<thead>
<tr>
<th>URL</th>
<th>Host Constraints</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.egjustice.org">www.egjustice.org</a></td>
<td>URL Contains user/register</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.egjustice.org">www.egjustice.org</a></td>
<td>URL Contains printmail</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.egjustice.org">www.egjustice.org</a></td>
<td>URL Contains /print/</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.egjustice.org">www.egjustice.org</a></td>
<td>URL Contains comment-form</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.egjustice.org">www.egjustice.org</a></td>
<td>URL Contains taxonomy</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.ehrco.org">www.ehrco.org</a></td>
<td>Add Rule</td>
<td>Add Rule</td>
</tr>
<tr>
<td><a href="http://www.elsam.or.id">www.elsam.or.id</a></td>
<td>Add Rule</td>
<td>70,000</td>
</tr>
<tr>
<td><a href="http://www.en.mosharka.org">www.en.mosharka.org</a></td>
<td>Add Rule</td>
<td>click me</td>
</tr>
<tr>
<td><a href="http://www.enoughproject.org">www.enoughproject.org</a></td>
<td>Add Rule</td>
<td>click me</td>
</tr>
<tr>
<td><a href="http://www.equalityhumanrights.com">www.equalityhumanrights.com</a></td>
<td>Add Rule</td>
<td>click me</td>
</tr>
<tr>
<td><a href="http://www.equalityhumanrights.com">www.equalityhumanrights.com</a></td>
<td>Reg Expr: *(col=yellow</td>
<td>col=black</td>
</tr>
<tr>
<td><a href="http://www.equalnighttrust.org">www.equalnighttrust.org</a></td>
<td>URL Contains email-a-friend</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.equalnighttrust.org">www.equalnighttrust.org</a></td>
<td>URL Contains &amp;q&amp;q</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>Add Rule</td>
<td>click me</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>URL Contains origen=fbk</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>URL Contains origen=tw</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>URL Contains tx_tt</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>URL Contains /enviar-esta-pagina/</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>URL Contains /ads/adlog</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>URL Contains https</td>
<td>Edit x</td>
</tr>
<tr>
<td><a href="http://www.es.amnesty.org">www.es.amnesty.org</a></td>
<td>URL Contains enviar-aquesta-pagina</td>
<td>Edit x</td>
</tr>
</tbody>
</table>

More host constraints for crawl scoping
<table>
<thead>
<tr>
<th>Host</th>
<th>URLs</th>
<th>Data</th>
<th>New URLs</th>
<th>New Data</th>
<th>Queued</th>
<th>Robots.txt Blocked</th>
<th>Out of Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.518.org">www.518.org</a></td>
<td>104,322</td>
<td>5.0 GB</td>
<td>104,322</td>
<td>5.0 GB</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><a href="http://www.equalityhumanrights.com">www.equalityhumanrights.com</a></td>
<td>96,129</td>
<td>2.6 GB</td>
<td>96,129</td>
<td>2.6 GB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><a href="http://www.ihr.org.uk">www.ihr.org.uk</a></td>
<td>20,270</td>
<td>1.3 GB</td>
<td>20,269</td>
<td>1.3 GB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eng.518.org</td>
<td>6,743</td>
<td>701 MB</td>
<td>6,743</td>
<td>701 MB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><a href="http://www.raipon.info">www.raipon.info</a></td>
<td>5,646</td>
<td>489 MB</td>
<td>5,646</td>
<td>489 MB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>518.org</td>
<td>3,691</td>
<td>118 MB</td>
<td>3,691</td>
<td>118 MB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><a href="http://www.hafsite.org">www.hafsite.org</a></td>
<td>977</td>
<td>61 MB</td>
<td>977</td>
<td>61 MB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><a href="http://www.ifhro.org">www.ifhro.org</a></td>
<td>723</td>
<td>108 MB</td>
<td>723</td>
<td>108 MB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><a href="http://www.youtube.com">www.youtube.com</a></td>
<td>236</td>
<td>11 MB</td>
<td>235</td>
<td>11 MB</td>
<td>0</td>
<td>4,145</td>
<td>0</td>
</tr>
<tr>
<td><a href="http://www.jpnuk.org.uk">www.jpnuk.org.uk</a></td>
<td>200</td>
<td>5.4 MB</td>
<td>200</td>
<td>5.4 MB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>desdecuba.com</td>
<td>136</td>
<td>233 KB</td>
<td>136</td>
<td>233 KB</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td><a href="http://www.gwangjuin.com">www.gwangjuin.com</a></td>
<td>96</td>
<td>8.2 MB</td>
<td>96</td>
<td>8.2 MB</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>gjdream.com</td>
<td>74</td>
<td>5.9 MB</td>
<td>74</td>
<td>5.9 MB</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><a href="http://www.desdecuba.com">www.desdecuba.com</a></td>
<td>71</td>
<td>3.6 MB</td>
<td>71</td>
<td>3.6 MB</td>
<td>0</td>
<td>0</td>
<td>434</td>
</tr>
</tbody>
</table>

Archive-It test crawl host report
Crawl scoping challenge: Website design generates multiple versions of all pages
http://www.equalityhumanrights.com/
Description and access for archived websites at CUL

- CLIO collection-level MARC records  
  (Human rights, Avery, Burke)

- CLIO site-level MARC records  
  (Human rights, Avery)

- Document-level MARC records  
  (selected longish Avery collection reports, pre-existing IRCR records)

- Archive-it.org metadata  
  (All thematic collections, DCMI, copied from MARC records if possible)

- Human Rights Web Archive portal on CUL website  
  (using metadata extracted from MARC records)
Avery Library historic preservation and urban planning web archive.

Title: Avery Library historic preservation and urban planning web archive.
Published: 2010-

Online Link(s): Full collection access via Archive-It.org

Restrictions: Browsing and full-text search of archived websites is available without restriction.

LC Subjects:
Atlantic Yards (Project)
Historic preservation--New York (State)--New York.
Historic preservation.
City planning--New York (State)--New York.
Historic districts--Conservation and restoration--New York (State)--New York.
Public spaces--New York (State)--New York.
Parks--New York (State)--New York.
Web archives.
New York (N.Y.)--Buildings, structures, etc.

Also Listed Under:
Avery Library.

Holdings Information:
Location (guide): Online
Call Number: Avery Library Historic Preservation and Urban Planning
Status: No information available

Other Subject Terms: Web archives

Summary: A growing collection of websites selected by the Avery Architectural and Fine Arts Library staff for web archiving preservation by the Columbia University Libraries' Web Resources Collection Program. Website captures began in 2010 and are ongoing. The collection's principal thematic focus is documenting the evolution of the built environment and public spaces through the interaction of historic preservation efforts and new development projects within urban planning debates. Selected websites are mostly published by non-profit groups or individuals based in the New York City area, including historic preservation groups, neighborhood associations, public policy organizations, parks conservancies, and both sponsors and critics of ongoing development projects. Websites on related themes from other
### Anti Caste Discrimination Alliance

**Author:** Anti Caste Discrimination Alliance.

**Title:** Anti Caste Discrimination Alliance [electronic resource].

**Published:** Derby : Anti Caste Discrimination Alliance

**Description:** 1 online resource.

**Online Link(s):**
- Current site
- Archived site, Oct. 2011-
- Archived site, Dec. 2010-Sep. 2011

**LC Subjects:**
- Anti Caste Discrimination Alliance.
- Caste-based discrimination--Great Britain.
- Caste--Great Britain.
- Discrimination--Great Britain.
- East Indians--Great Britain.

**Holdings Information:**

<table>
<thead>
<tr>
<th>Location (guide)</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call Number</td>
<td>ERESOURCES</td>
</tr>
<tr>
<td>Status</td>
<td>No information available</td>
</tr>
</tbody>
</table>

**Other Subject Terms:** Non-governmental organizations

**Summary:** "ACDA is an umbrella body for a number of groups and organisations who are working to eliminate Caste Discrimination in the UK and more widely."

**Other Title:** ACDA

**Notes:** Viewed on Oct. 18, 2010.

**Material Type:** Other

**Preserved by Columbia University Libraries' Web Resources Collection Program.**

**Bookmark this page as:** http://clio.cul.columbia.edu:7018/vwebv/holdingsInfo?bibId=8067914

**CLIO public view of site-level MARC record**
<table>
<thead>
<tr>
<th><strong>Portal Field</strong></th>
<th><strong>MARC Tag</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>245</td>
</tr>
<tr>
<td>Creator</td>
<td>1XX/7XX</td>
</tr>
<tr>
<td>Organization Type</td>
<td>653 0</td>
</tr>
<tr>
<td>Organization Based In</td>
<td>&quot;Country&quot;</td>
</tr>
<tr>
<td>Geographic Focus</td>
<td>043 or</td>
</tr>
<tr>
<td></td>
<td>965 = &quot;Global focus&quot;</td>
</tr>
<tr>
<td>Subjects</td>
<td>650 ($a, $x)</td>
</tr>
<tr>
<td>Summary</td>
<td>520</td>
</tr>
<tr>
<td>Languages</td>
<td>&quot;Language&quot; and 041</td>
</tr>
<tr>
<td>Seed URLs</td>
<td>856 40</td>
</tr>
<tr>
<td>Archived URLs</td>
<td>856 41</td>
</tr>
<tr>
<td>Nonpublic identifier</td>
<td>“965hrportal”</td>
</tr>
</tbody>
</table>

Mapping MARC data for portal facets
Human Rights Web Archive portal

Features

- CUL customized access portal to Human rights web archive
- Browse archived websites by Title, URL, Subject, Place (Geographic focus), Language, using facet data pulled in from CLIO MARC records
- Search website descriptions (i.e. MARC metadata)
- Search page full text (based on local SOLR indexing of .WARC file data downloaded from Internet Archive and extracted into a php database)
- "Nominate a site" forms for site owners and general public

Development features soon to be added to public portal

- Search extension -- search query extended (by user's opt-in) via API to any of a list of external resources related to web archives and/or human rights
- Search expansion -- If query term matches any term on pre-constructed list of names of people groups and their synonyms, users offered the option of expanding their search to include the relevant synonyms
The Human Rights Web Archive @ Columbia University is a searchable collection of archived copies of human rights websites created by non-governmental organizations, national human rights institutions, tribunals and individuals. Collecting began in 2008 and has been ongoing for active websites. New websites are added to the collection regularly.

The HRWA is an initiative of the Center for Human Rights Documentation & Research and is a key focus of the Columbia University Libraries' Web Resources Collection Program. The HRWA was made possible by generous support from the Andrew W. Mellon Foundation. Learn More »

http://hrwa.cul.columbia.edu
<table>
<thead>
<tr>
<th>Titles</th>
<th>Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;İnsan Hüquqları XXI asr-Azərbaycan&quot; Fondu</td>
<td>Human Rights Council of Australia</td>
</tr>
<tr>
<td>5-18 Kinyõm Chaedan</td>
<td>Human Rights Education Institute of Burma</td>
</tr>
<tr>
<td>ABONG--Associação Brasileira de Organizações Não Governamentais</td>
<td>Human Rights First</td>
</tr>
<tr>
<td>Abuelas de Plaza de Mayo</td>
<td>Human Rights First Society</td>
</tr>
<tr>
<td>Access to Justice</td>
<td>Human Rights Law Network</td>
</tr>
<tr>
<td>Ação Brasileira pela Nutrição e Direitos Humanos</td>
<td>Human Rights Trust of Southern Africa</td>
</tr>
<tr>
<td>Adalah</td>
<td>Human Rights Watch</td>
</tr>
<tr>
<td>Advocacy Forum--Nepal</td>
<td>humanrights.ge, web portal on human rights in Georgia</td>
</tr>
<tr>
<td>The Advocates for Human Rights</td>
<td>I-India</td>
</tr>
<tr>
<td>AFAPREDESA</td>
<td>IDASA</td>
</tr>
<tr>
<td>Afghanistan Independent Human Rights Commission</td>
<td>Independent Commission for Human Rights</td>
</tr>
<tr>
<td>Africa Internally Displaced Persons Voice</td>
<td>Indeso mujer</td>
</tr>
<tr>
<td>African Centre for Democracy and Human Rights Studies</td>
<td>INDICT</td>
</tr>
<tr>
<td>African Commission on Human and Peoples’ Rights</td>
<td>Indigenous Peoples Council on Biocolonialism</td>
</tr>
<tr>
<td>Al-Dameer Association for Human Rights</td>
<td>Indigenous Peoples’ Center for Documentation, Research and Information</td>
</tr>
<tr>
<td>Al-Haq</td>
<td>Indignación</td>
</tr>
<tr>
<td>Albanian Center for Human Rights</td>
<td>Informatsionno-analiticheskii tsentr “Sova”</td>
</tr>
</tbody>
</table>
### HRWA portal: Browse by subject (sortable by count)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human rights (General)</td>
<td>425</td>
</tr>
<tr>
<td>Civil rights</td>
<td>67</td>
</tr>
<tr>
<td>Democracy</td>
<td>38</td>
</tr>
<tr>
<td>Women's rights</td>
<td>29</td>
</tr>
<tr>
<td>Ombudspersons</td>
<td>22</td>
</tr>
<tr>
<td>Civil society</td>
<td>15</td>
</tr>
<tr>
<td>Transitional justice</td>
<td>15</td>
</tr>
<tr>
<td>Torture</td>
<td>14</td>
</tr>
<tr>
<td>Indigenous peoples</td>
<td>14</td>
</tr>
<tr>
<td>Truth commissions</td>
<td>12</td>
</tr>
<tr>
<td>Refugees</td>
<td>11</td>
</tr>
<tr>
<td>Rule of law</td>
<td>11</td>
</tr>
<tr>
<td>Discrimination</td>
<td>9</td>
</tr>
<tr>
<td>Disappeared persons</td>
<td>9</td>
</tr>
<tr>
<td>Palestinian Arabs</td>
<td>9</td>
</tr>
<tr>
<td>Minorities</td>
<td>8</td>
</tr>
<tr>
<td>Freedom of the press</td>
<td>7</td>
</tr>
<tr>
<td>Impunity</td>
<td>7</td>
</tr>
<tr>
<td>Reconciliation</td>
<td>7</td>
</tr>
<tr>
<td>Legal aid</td>
<td>7</td>
</tr>
<tr>
<td>Political prisoners</td>
<td>7</td>
</tr>
<tr>
<td>Peace-building</td>
<td>6</td>
</tr>
<tr>
<td>Genocide</td>
<td>6</td>
</tr>
<tr>
<td>Gun control - International cooperation</td>
<td>1</td>
</tr>
<tr>
<td>Children and war</td>
<td>1</td>
</tr>
<tr>
<td>Hazardous wastes -- Environmental aspects</td>
<td>1</td>
</tr>
<tr>
<td>Health -- Social aspects</td>
<td>1</td>
</tr>
<tr>
<td>Hindu -- Diaspora</td>
<td>1</td>
</tr>
<tr>
<td>Hinduism</td>
<td>1</td>
</tr>
<tr>
<td>Hindus</td>
<td>1</td>
</tr>
<tr>
<td>Historians -- Political activity</td>
<td>1</td>
</tr>
<tr>
<td>HIV infections</td>
<td>1</td>
</tr>
<tr>
<td>Homosexuality</td>
<td>1</td>
</tr>
<tr>
<td>African diaspora</td>
<td>1</td>
</tr>
<tr>
<td>Housing</td>
<td>1</td>
</tr>
<tr>
<td>Church and state</td>
<td>1</td>
</tr>
<tr>
<td>Human rights -- Environmental aspects</td>
<td>1</td>
</tr>
<tr>
<td>Human rights -- Government policy</td>
<td>1</td>
</tr>
<tr>
<td>Human rights -- Health aspects</td>
<td>1</td>
</tr>
<tr>
<td>Human rights -- History</td>
<td>1</td>
</tr>
<tr>
<td>Human rights -- Study and teaching</td>
<td>1</td>
</tr>
<tr>
<td>Human rights and globalization</td>
<td>1</td>
</tr>
<tr>
<td>Arms transfers - Prohibition</td>
<td>1</td>
</tr>
<tr>
<td>Africans -- Civil rights</td>
<td>1</td>
</tr>
<tr>
<td><strong>Alternate Title</strong></td>
<td>ACDA</td>
</tr>
<tr>
<td>--------------------</td>
<td>------</td>
</tr>
<tr>
<td><strong>Creator</strong></td>
<td>Anti Caste Discrimination Alliance</td>
</tr>
<tr>
<td><strong>Organization Type</strong></td>
<td>Non-governmental organizations</td>
</tr>
<tr>
<td><strong>Organization Based In</strong></td>
<td>United Kingdom</td>
</tr>
<tr>
<td><strong>Geographic Focus</strong></td>
<td>Great Britain</td>
</tr>
<tr>
<td><strong>Subjects</strong></td>
<td>Caste-based discrimination; Caste; Discrimination; East Indians</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>&quot;ACDA is an umbrella body for a number groups and organisations who are working to eliminate Caste Discrimination in the UK and more widely.&quot;</td>
</tr>
<tr>
<td><strong>Languages</strong></td>
<td>English</td>
</tr>
</tbody>
</table>

HRWA portal : Site details view, with links to crawl calendar page(s)
HRWA portal: Search website descriptions (metadata)
HRWA portal : Website description search results
Displaying page 1 for about 1,100,000 archived web page results (0.547 seconds) for: trafficking

Trafficking
http://www.ihrc.ie/download/ppt/paper200909_demand_immigrantcouncil.ppt
April 16, 2012 at 8:06PM UTC
Trafficking International Human Rights Law and State Practice Sex Trafficking and Prostitution: The Dilemma of Demand Tuesday, 23rd June 2009 Council of Europe Convention on Action against Trafficking in Human Beings Article 6 – Measures to discourage the demand To discourage the demand that fosters all forms of exploitation of persons, especially women and children, that leads to trafficking, each Party shall adopt or strengthen legislative, administrative, educational, social, cultural or Size: 115 KB - Presentation

Trafficking
http://worecnepal.org/print/43
Nov 7, 2011 at 7:33PM UTC
Trafficking Published on WOREC Nepal (http://worecnepal.org) Home > Programs > Trafficking Trafficking Definition of Trafficking Trafficking in persons means the recruitment, transportation, purchase, sale, transfer, harbouring or receipt of persons: by threat or use of violence, abduction, force, fraud, deception or coercion (including the abuse of authority) or debt bondage, for the purpose of placing or holding such person, whether for pay or not, in forced labour or slavery-like. Size: 12.3 KB - HTML

Trafficking
May 14, 2009 at 4:14PM UTC
Trafficking National Human Right Commission - Combating Trafficking In Women and Children - Padma Mathema National Rapporteur Introduction The problem of trafficking in women and children is an issue of serious concern over the years at the national and international level. As such, it is the decades old serious in - human crime depriving total dignity and freedom and is the total violation of human rights. On one hand, the effort of the 21st century world is in the process of improving the quality of life of Size: 63.4 KB - PDF

Other Relevant Captures
Sep 2, 2010 at 9:17PM UTC (63.4 KB)

Trafficking
http://worecnepal.org/ne/print/43
Nov 7, 2011 at 7:47PM UTC
Trafficking Published on WOREC Nepal (http://worecnepal.org) Home > Programs > Trafficking Trafficking Definition of Trafficking Trafficking in persons means the recruitment, transportation, purchase, sale, transfer, harbouring or receipt of persons: by threat or use of violence, abduction, force, fraud, deception or coercion (including the abuse of authority) or debt bondage, for the purpose of placing or holding such person, whether for pay or not, in forced labour or slavery...
Size: 12.3 KB - HTML
HRWA trial feature: search extension
You just searched in HRWA for: death penalty

Try your search in one of these related resources. (Search results will display in a new window.)

**Archive-It**
All web archives created by Archive-It partner institutions, including Columbia University.

**HuriSearch**
Custom live web search engine for over 5,000 human rights websites powered by Huridocs.

**ArchiveGrid**
Database of nearly 2,000,000 archival collection descriptions.

**Office of the High Commissioner for Human Rights, United Nations**
Documents, news, and publications on the OHCHR website.

**Universal Human Rights Index, maintained by the OHCHR**
Information issued by the UN's international human rights mechanisms: Treaty Bodies, Special Procedures, and the Universal Periodic Review.
HRWA trial feature: Search expansion (include synonyms of names of peoples)
HRWA trial feature: Search expansion (include synonyms of names of peoples)
Memento wants to make it as straightforward to access the Web of the past as it is to access the current Web.

If you know the URI of a Web resource, the technical framework proposed by Memento allows you to see a version of that resource as it existed at some date in the past, by entering that URI in your browser like you always do and by specifying the desired date in a browser plug-in. Or you can actually browse the Web of the past by selecting a date and clicking away. Whatever you land upon will be versions of Web resources as they were around the selected date. Obviously, this will only work if previous versions are available somewhere on the Web. But if they are, and if they are on servers that support the Memento framework, you will get to them.

[Formal technical specifications] detailing the Memento framework are available. But if you want a more gentle entry point to gain an understanding of how Memento is trying to change the Web by adding a time dimension to its most common protocol, HTTP, check out the [Introduction to Memento].

If you are interested in establishing a Web with a memory, please join the [Memento Development Group]. And if you want to get involved in the technical discussion of the [Memento specification], you probably should subscribe to the [IETF HTTP list] or to the [W3C TAG list].

Our Memento ideas have attracted quite some attention since we first shared them in November 2009. And a lot has happened since then. Stay up-to-date by checking in on Memento news here.

[News about Memento]

A command line Memento client, named mcurl (after curl), has been released by Ahmed AlSum, PhD student of Michael L. Nelson of the Computer Science Department of Old Dominion University. This client is available via [http://www.cs.odu.edu/~aalsum/projects/mcurl.html](http://www.cs.odu.edu/~aalsum/projects/mcurl.html) and is compatible with version 07 of the Memento Internet Draft, which is available at [http://www.ietf.org/id/draft-vandesompel-mem](http://www.ietf.org/id/draft-vandesompel-mem).
Installing the bookmarklet...
To install the Memento look-up service in your browser, drag this link to your bookmarks bar:

Find Mementos

Then, while browsing, if you hit a 404 Not Found error, or are just interested in knowing the archival history of your current web page, you can click the bookmarklet and you will go directly to the Mementos service results for that URL.

For more information, see the Mementos homepage.

This web interface uses the Memento aggregate TimeGate hosted by lanl.gov.
For more information on Memento, see www.mementoweb.org.

http://www.webarchive.org.uk/mementos/search
Thanks!

Any questions, or comments?

Contact me at at2186@columbia.edu