Archive-It:

Archiving and Preserving Born Digital Content

Molly Bragg
Partner Specialist
Internet Archive

About Internet Archive

- Non profit founded in 1996 by Brewster Kahle
- Universal access to human knowledge
- Officially designated a library by the state of California (2007)
- Built on open source software and dedicated to open source principles
- Current archive is 150 billion pages
- Largest publicly accessible web archive: *www.archive.org*

How do we collect it?

Open Source Technology
primarily developed by Internet Archive and IIPC

- **Heritrix**: web crawler - crawls and captures pages
- **Wayback Machine**: access tool for rendering and viewing pages. Displays archived web pages--surf the web as it was.
- **NutchWAX**: Open source search engine. Standard full-text search
- **WARC File**: archival file format used for preservation – ISO standard

Archive-It

- Web based application that allows users to create, manage and preserve collections of born digital content.

- Annual subscription service, includes hosting, access and storage

- Partners do not need significant technical infrastructure or personnel resources

- Functions include: harvesting, scoping, full text search, cataloging with metadata, reports and analysis of collections

www.archive-it.org

Archive-It Partners

First deployed in January 2006
Current total: 102 partners

- 39% University and Public Libraries
- 30% State Archives and Libraries
- 10% High Schools
- 10% Non Government Non Profits
- 5% National Libraries
- 4% Federal Institutions
- 2% Museums

- http://www.archive-it.org/public/partners

Access to Born Digital Content

Access = Use = Funding

- Various ways to access collections online:
  - Private web application with login/password
  - Archive-It public website
  - Partners website: landing pages with institutions' layout, look and feel
  - Restricted and private access options available

Archiving the internet for future generations
Collect it, manage it, search it...ARCHIVE-IT

ARCHIVE-IT

| Partners | FAQ | About Us | Press Room | Contact Us | Partner Login |

Enter Search Terms Here...   Select an Institution ▼

...Or Search All Collections ▼ **GO** Advanced Search

WELCOME TO ARCHIVE-IT

▶ **Attend a live online demo: March 24 11:30AM PDT or April 14 11:30AM PDT**
  -or-
▶ **Watch a 20 minute pre-recorded online demo now**

▶ As of March 24, 2009: Archive-It has collected **816,993,562** URLs for **765** public collections!

Archive-It, a subscription service from the Internet Archive, allows institutions to build and preserve collections of born digital content. Through our user-friendly web application, Archive-It partners can harvest, catalog, manage, and browse their archived collections. Collections are hosted at the Internet Archive data center and are accessible to the public with full-text search.

Browse Our Partners

State Archives & Libraries
Choose a partner

Colleges & Universities
Choose a partner

Museums & Public Libraries
Choose a partner ▼

National Institutions
Choose a partner ▼

K-12 Project Schools
Choose a partner

Other Institutions
Choose a partner ▼

Featured Collections

Latin American Government Documents Archive, LAGDA

The Latin American Government Documents Archive (LAGDA) seeks to preserve and facilitate access to a wide range of ministerial and presidential documents from 18 Latin American and Caribbean countries.

National Library of Australia: Papua New Guinea Government and Research Websites

Selected Papua New Guinean governmental and significant research institute web sites archived from 2008.

Browse Our Collections

▶ **Public Collections A-Z**

▶ **Arts & Humanities**

▶ **Computers & Technology**

▶ **Government**

▶ **Spontaneous Events**

▶ **Politics & Elections**

▶ **Science & Health**

▶ **Society & Culture**

▶ **Universities & Libraries**

Arizona State Agency Publications
Contributed by Arizona State Library, Archives and Public Records

ARIZONA STATE AGENCY PUBLICATIONS

Click here to view this collection

Contributed by Arizona State Library, Archives and Public Records

The Arizona State Agency Publications collection has been contributed by the Arizona State Library, Archives and Public Records' Law and Research Library. Digitized publications produced by and for Arizona state agencies are available from this collection, which is updated on a continuing basis. Also included are a few reports that were not published for or by a state agency. These were received by us from the Legislature for inclusion in the collection.

Popular searches:

- Annual Reports
- House Received Reports
- Legislative Study Committee Reports
- Senate Received Reports
- Territorial Documents

Also try our Executive Orders and Attorney General Opinions collections!

If you haven't found what you are looking for, please try the Library and Archives' online catalog, visit the library, or ask a librarian!

Search OLD versions of Arizona state agency websites:

We archive information published on state agency websites on a regular basis. Enter your search terms below to search this older content. Your search results will be loaded into a new page. Use your browser's back button to return to this page.

[   ] Go

Arizona State Library, Archives and Public Records
Law and Research Library
Suite 300
1700 W Washington Ave
Phoenix, AZ 85007
(602) 926-3870

What is compelling about archived web content?

- "At risk" content needs to be preserved before it is lost

- More primary source information is only available in born-digital format

- Diverse range of content included in one location (website)

- Need to document history from multiple perspectives for future generations

Archive-It Application

Internet Archive - Web Archive
Partner Since February 2006

Partner Home

Current Subscription (started on 01-Jan)

Documents Crawled:	6,685,357
Subscription Document Budget:	12,000,000
Budget Used:	55.71%
Data Archived:	238.83 GB
Maximum Seed Count	300
Total Active Seeds:	65

XML

All Subscription Periods

Documents Crawled:	42,452,831
Data Archived:	2,083.24 GB

Getting Started
Create New Collection

Active Collections	Last Completed Crawl	Next Scheduled Crawl
Net videos	June 11, 2009 11:47:43 PM GMT	September 8, 2009 11:10:41 PM GMT
Online Books	March 27, 2009 3:44:02 AM GMT	June 24, 2009 3:35:26 AM GMT
San Francisco 1906 Quake	April 21, 2006 6:53:54 PM GMT	

San Francisco 1906 Quake   Collection Management

Created: internetarchive Apr 11, 2006 3:46:02 AM
Updated: internetarchive Jun 12, 2007 1:21:40 PM   **XML** 💡 [Activate] [Deactivate] [Mark Dormant]

Collection Management   ## Seed Management

Add Seeds   by seed state: by crawl frequency:
Modify Crawl Scope   All (52)   Daily (0)   Quarterly (0)
Edit Collection Metadata   Active (52)   Weekly (0)   Semiannual (0)
View Reports   Inactive (0)   Monthly (0)   Annual (0)
                                                               One-Time (52)

Crawling Activity

Frequency	Last Completed Crawl	Next Scheduled Crawl	
One-Time			**Start Crawl Now**

ARCHIVE-IT

Scope-It

Scope-It is a discovery tool designed to help Archive-It partners be more selective about the content captured for a collection.

To get started, please select one of the options below.

| Crawl Seeds and Import | Specify a list of URLs to be crawled and then imported into the Scope-It Crawl Explorer for analysis. |

| Import Completed Crawl | Select a crawl from any collection to be imported into the Scope-It Crawl Explorer for analysis. |

| Monitor Scope-It Crawls and Imports | Back to Home |

Home   Collections   Crawls   Reports   Access   Help

Access   Wayback

Search   Wayback

Specify an exact URL [   ] Go

Seeds in Collection [San Francisco 1906 Quake   ▲▼]

http://1906centennial.org/
http://1906expo.com/
http://bancroft.berkeley.edu/
http://bancroft.berkeley.edu/collections/earthquakeandfire/
http://bancroft.berkeley.edu/collections/earthquakeandfire/splash.html
http://earthquake.usgs.gov/
http://earthquake.usgs.gov/regional/nca/1906/
http://earthquake.usgs.gov/regional/nca/virtualtour/
http://exploratorium.edu/
http://kron4.com/
http://memory.loc.gov/ammem/papr/sfhome.html
http://news.yahoo.com/s/ap/20060418/ap_on_sc/earthquake_study
http://photo.ucr.edu/california/projects/sf1906/
http://quake.wr.usgs.gov/
http://quake.wr.usgs.gov/info/1906/
http://seismo.berkeley.edu/1906/
http://sfgate.com/
http://sfgate.com/greatquake
http://sfhistory.org/

Access

Search   Wayback

Find Documents:

with all of the words   ossetia

In collection   Georgia and Russia Conflict, August 2008 ⬍

Search   Help   [Advanced Search]

Hits 1 - 10 (out of a total of approximately 419,707 matches)

Global Voices Online » South Ossetia Crisis 2008

... search for itself by Alexander Rondeli, South Ossetia: the avoidable tragedy by Thomas de ... End to Hostilities in South Ossetia - From the Pulitzer Center on Crisis ... on the Untold Stories blog Photos - South Ossetia Flickr Set by Joshua Kucera Online activism Avaaz ... Editor, Onnik Krikorian . Map of South Ossetia's location in Europe , South Ossetia ... 2008 AIDS 2008 Conference Beijing Oympics 2008 South Ossetia Crisis 2008 Other Sections About For Bloggers ...

http://globalvoicesonline.org/specialcoverage/south-ossetia-crisis-2008/ - 638306-01-31 05:43:59 PM GMT - - other versions - more from globalvoicesonline.org

Осетия-Алания :: Северная Осетия : Владикавказ :: Южная Осетия : Цхинвал

... 30.08.2008 Откровения грузинских солдат [182] Редакция Ossetia.ru перевела откровения грузинских солдат и резервистов, воевавших ... Google. Перевод историй даётся без изменений, практически дословный. Ossetia.ru не комментирует содержание текстов оставляя это читателям ... большую занятость президента Южной Осетии Эдуарда Кокойты, редакции Ossetia.ru удалось взять у него блиц-интервью. Прямо ...

How Partners Use Archive-It

Stanford University, Islamic and Middle Eastern Collection

Purpose: harvest and preserve Iranian Blogs

- Archiving over 300 blogs written by and for Iran and the Iranian people

- Also includes coverage of current Iranian elections

- Partner since February 2008

- 16 million URLs, 1.4 terabytes of data

روزنگار

حرف دل با اهل دل

از اکنون در حال نقشه کشیدن و برنامه ریزی برای تعطیلات نوروز 86 هستم . هنوز زمستان از راه نرسیده، سرم سودای بهار را می پروراند . بهاری همراه با سفر به گوشه و کنار ایران و اگر مقدماتی مهیا شود سفری خواهم داشت به یکی از کشورهای همسایه تا شخصیتی را ملاقات کنم که تنها راه ارتباط ما سیمهای تلفن و صفحه جادویی تلویزیون است . اسمی از این نازنین مرد نخواهم برد تا مبادا هم برای من و هم برای ایشان مشکلی پیش آید چون وی به میمنت و مبارکی وجود حضرات آیات ا امکان سفر به ایران را ندارد . هفته گذشته که مژده چنین دیداری را از زبان این هنرمند برجسته شنیدم شور و حال عجیبی دارم . گویی دنیا را به من داده اند . بگذارید به لطف روزنگار این سفر و دیدار انجام شود آنگاه به شما خواهم گفت این همه شوق برای دیدن کیست .

اما این یک روی سکه بود . آن روی دیگر سکه چندان نقش خوشی ندارد . جمعه پیش کوروش یغمایی بودم . هر از چند گامی با هم دیداری داریم و با هم به گفتگو می نشینیم . حدود 9 ماه است که منتظر صدور مجوز آلبوم جدیدش (ملک جمشید) است و بنا به دلایلی که امکان مطرح کردن آن در این موقعیت زمانی نیست، نه خبری از صدور مجوز است و نه چیز دیگر . همین نا مهربانی ها او را مجبور کرده تا خود را برای جلای وطن و پیوستن به محفل غربت نشینان آماده کند .

صفحه اصلی
نوشته ها
سخنرانی و مصاحبه
زندگینامه
آلبوم عکس
آثار هنری
دیدگاه های شما
ارتباط با من

پایگاه های مرتبط

قلم نیوز

کلمه

www.sepidedam.com

سپیده دم

چرا سبز ؟

رنگ سبز ، به عنوان نماد فعالیتهای انتخاباتی در حال گسترش است . در مورد این رنگ نگاهها و باورهای مختلفی وجود داشته و اکنون طیفها و گروههای متعددی بنا بر خواست ، باور ، یا موجودیت خود در تمام جهان رنگ سبز را به عنوان نماد برگزیده اند . برپایه دلشغولیهای هنریم می دانم که رنگها عوارضی درحال گذشتن نیستند . آنها از جنس واقعیتند تاثیر می گذارند و از وجود و عواطف انسانها متاثر می شوند . قرنها پیش و برای همیشه تاریخ ، ائمه اطهار علیهم السلام این رنگ را انتخاب کردند . برای آن بزرگواران حتما انتخاب رنگ سبز دلایل هستی شناسانه ای داشته و اتفاقی نبوده است .

قرآن کریم لباس بهشتیان را سبز دانسته است و براساس بعضی احادیث ، سبز رنگ ردابیست که پیامبر (ص) در قیامت می پوشد . سبز رنگ نماد اهل بیت و پرچم امام علی (ع) است و در هجمه ظلم و بی عدالتی یادآور امید و انتظار ما برای ظهور حضرت مهدی (ع) است . ما ایرانیان نیز شادی هر شروع نیکی را باسبز نشان می دهیم و دعا می کنیم که زندگی کسانی که عزیزمی داریم سبز باشد .
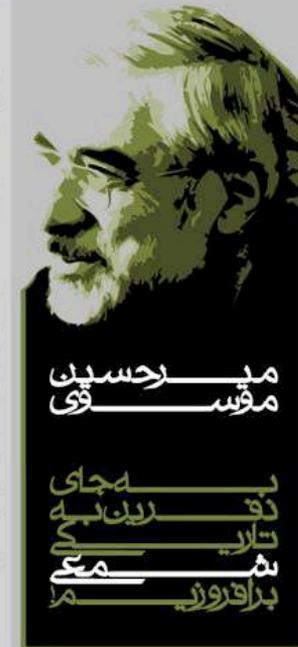
همه دوستان می دانند که امروزه بخش بزرگی از مبادلات اطلاعات و اتفاقات برای کل و عموم جامعه بر عهده تلوزیون و رادیوست اما ظاهرا اکنون این رسانه ها تنها متعهد به گزینش اطلاعات به نفع خواست و مطالبات گروه ها و افراد موردنظر و انتقال گاه غیر امانتدارانه اطلاعات از جانب دیگر افراد فعالند . در نبود و یا حضور غیر عادلانه و غیر دوستانه رسانه ملی این رنگ بهترین رسانه برای ماست تا مظلومیت خود را به گوش همگان برسانیم و البته برای ما به واسطه آن سابقه معنایی که پیشتر گفته شد رنگ سبز به معنای متعهد بودن و باورداشتن به ارزشهای مذهبی نیز هست . این رنگ را برگزیده شد که با شعارهایی که قبولشان داریم منطبق است ، پیام آور شادیست .

خدایا تنها تورا می پرستم و از تو کمک می خواهم و با خود زمزمه می کنم

برگ درختان سبز در نظر هوشیار   هر ورقش دفتریست معرفت کردگار

نظرات شما »

[آرشیو]

میر حسین موسوی

به جای دیر این به تاریکی شمعی برافروزیم

آمار مراجعات

مهمان : 48

Virginia Tech University

Purpose: capture an event as it unfolds on the web and changes rapidly

- Quick set-up and archive on demand
- University sites, news sites, blogs
- Crisis, Tragedy and Preservation Consortium
- Northern Illinois University shooting (Feb 08)
- 5.3 million URLs, 330 gigabytes of data

THE APRIL 16 ARCHIVE
COLLECTING AND PRESERVING THE STORIES OF THE VIRGINIA TECH TRAGEDY

Featured Item

A Day of Remembrance Vigil

Candlelight vigil held on the drillfield on 4/16/08, A Day of
Remembrace. Burruss Hall seen in the background. Photo by
Kevin...

Recently Added

- A year later (written on the first anniversary by an Alum)
- Before the permanent memorial
- Norris classes moved
- Bear Hugs
- Drillfield tent
- No More Media
- Memorial After Rememberance
- Flags at Blacksburg Baptist Church

Electronic Literature Organization

Purpose: archive born digital literature

- Poems and stories that are generated by computers, either interactively or based on parameters given at the beginning

- Collect individual works, collections/journals, and critical opinion

- Archive-It Partner since July 2007

- 5.6 million URLs, 340 gb of data

drag to rotate

2009 – 2010 Programs

- K12 Web Archiving Program
  - 9 schools 2008 – 2009
  - www.archive-it.org/k12/
  - Applications for 2009 -2010 program begin mid July: www.loc.gov/teachers

- Spanish User Interface
  - Global Spanish speaking partners
  - US Hispanic Population

Archivando Internet para las futuras generaciones
Colecciónela, manéjela, búsquela...ARCHIVE-IT

ARCHIVE-IT

| Socios | FAQ | Acerca de Nosotros | Sala de Prensa | Contáctenos | Ingreso de Socios |

Ingrese aquí los términos a busc | Seleccione una Institución ⬍ | ...O busque en Todas las Colecciones ⬍ IR
Búsqueda Avanzada

BIENVENIDO A ARCHIVE-IT

▶ Asista a una demostración en línea: July 14 11:30AM PDT o July 28 11:30AM PDT
-o-
▶ Vea ahora, en línea, demostraciones pre-grabadas

▶ Hasta June 25, 2009:
Archive-It ha cosechado
1,025,363,829 URLs para
871 colecciones públicas!

Archive-It, un servicio de suscripción del Internet Archive, permite a las instituciones suscritas confeccionar y preservar colecciones de contenido cosechado de la Web. Usando nuestra sencilla aplicación, los socios de Archive-It pueden cosechar, catalogar, manejar, y consultar sus colecciones archivadas. Las colecciones son alojadas en el centro de datos del Internet Archive y pueden ser consultadas por el público en general en la Web por medio de búsquedas de texto completo.

Consultar la lista de nuestros socios

Archivos y bibliotecas de los estados

Selecciona un socio

Colegios y universidades

Selecciona un socio

Museos y bibliotecas públicas

Selecciona un socio ⬍

Instituciones nacionales

Colecciones Destacadas

Biblioteca Nacional de Chile: DIBAM

Sitios web pertenecientes a la Dirección de Bibliotecas, Archivos y Museos de Chile. Esta colección de la Biblioteca Nacional de Chile, en Santiago, resalta las actuales actividades de archivo y preservación en América Latina en bibliotecas, museos y archivos de Chile.

English Español Português

Archivo de Documentos Gubernamentales de América Latina
Acerca del Archivo

Consultar la lista de nuestras colecciones

▶ Colecciones de A hasta Z

▶ Artes y Humanidades

▶ Computadoras y Tecnología

▶ Gobierno

▶ Eventos Espontáneos

K-12 Web Archiving

ABOUT THE PROGRAM

If you were a student which websites would you want to save for future generations? This is the challenge we posed to students and their teachers. For the 2008/2009 school year, Internet Archive, the Library of Congress and California Digital Library collaborated on a program that explores archiving the Web from the perspective of students in elementary, middle and high school.

We hope that stimulating students to think about history in the context of their own lives will provide them an opportunity to actively engage in selecting the matter of history in the future, and help students begin to grasp the tremendous challenges presented by a world in which information can be both generated and annihilated in a heartbeat.

Students at Moran Middle School

Using the Archive-It service, students from ten different schools selected born digital content from the Web to create "time capsules" to represent their world. By allowing students to identify sites that will be preserved for the long-term, the program gives teens and younger students a chance to identify and document their cultural history and the world that's important to them. Unlike time capsules of tangible objects, which usually remain hidden for decades or centuries, the resulting Web collections are immediately visible and publicly accessible here, with full text search for study and analysis.

For the 2009/2010 school year we hope to broaden the program's outreach to additional schools around the country. To get involved and/or learn more, please send us your information through this request form. Applications will be sent out mid to late July.

WHAT TEACHERS AND STUDENTS ARE SAYING

"My students view internet documents differently now. They are more aware of the continually changing nature of the Internet and the impact that has on history. They felt empowered to speak for their generation about what is important to them, what they value and how they communicate."

- Jayne Staniforth, Extended Learning Program teacher, Ames Middle School

"I think this is a very cool project for people our age to be doing. It really involves us in something important and these archives will be around forever."

www.archive-it.org/k12/

Thank you!

Molly Bragg

Partner Specialist

415.561.6799, ext. 6

mbragg@archive.org

Kristine Hanna

Director, Web Archiving Services

415.561.6799m ext. 5

kristine@archive.org

www.archive-it.org