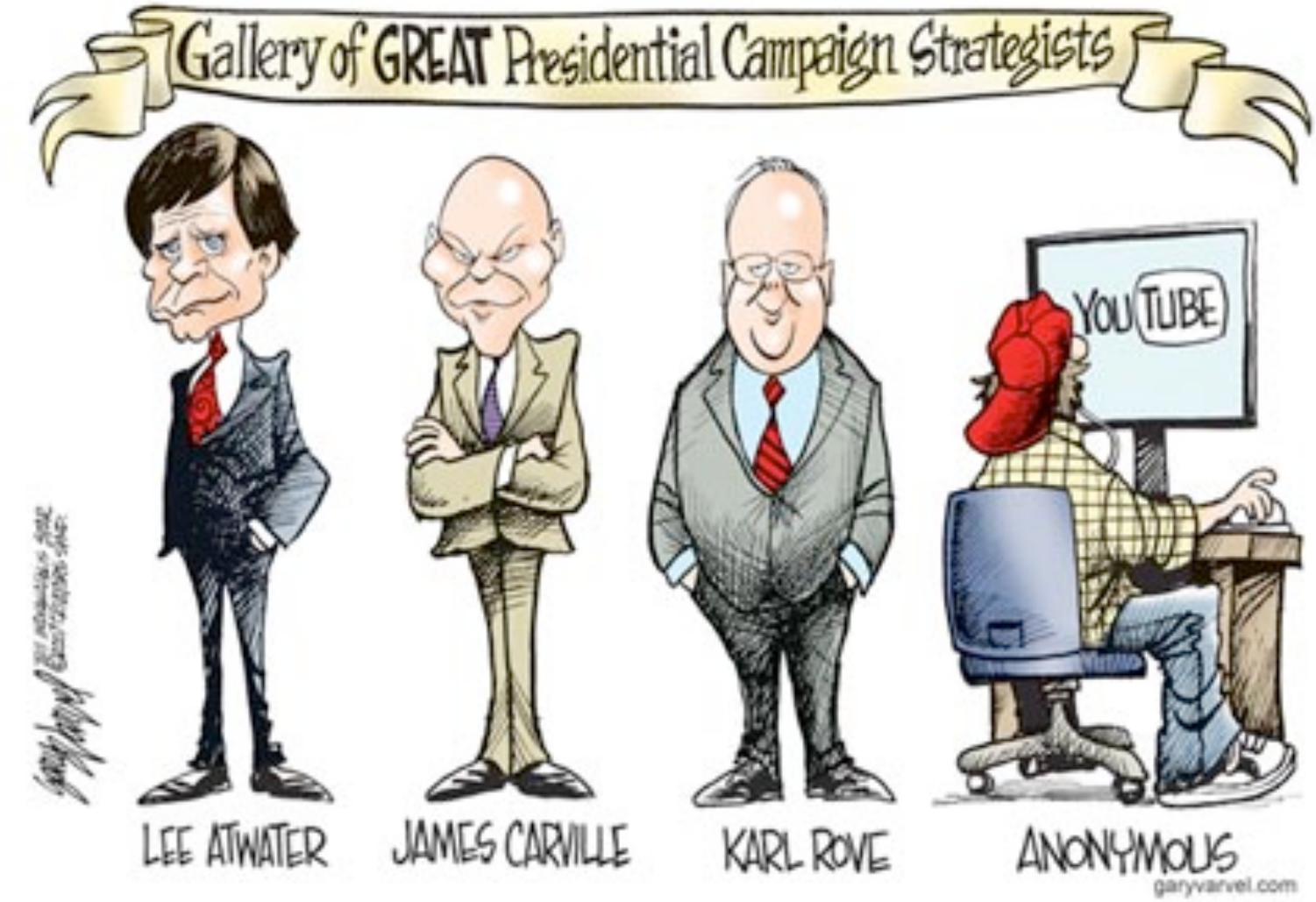
VidArch
Telling the Whole Story: Augmenting Archival Documentation Strategies

Helen R. Tibbo
School of Information and Library Science
University of North Carolina at Chapel Hill
tibbo@ils.unc.edu

YouTube and the 2008 Presidential Elections

Video Harvesting from YouTube for Election 2008



VidArch

- Funded initially by NSF
  - "Preserving Video Objects and Context: A Demonstration Project"
  - Supported by the National Science Foundation #IIS 455970 DigArch Program
- Current grant funded by LOC – NDIIPP
  - Extending archival documentation strategies
  - Partners: SDSC and Internet Archive

Today's Presentation

- Goals of the VidArch Project
- Archival Appraisal
- Archival Documentation Strategies
- You Tube and Similar Collections
- Context Miner
- Collecting Scenario
- Next Steps

Goals of this Project

- Develop systematic approaches to selection of videos found on the Web for archival collections so that curators can more easily collect and preserve materials in a wide array of formats from text to audio to video.
- Develop tools to assist with collecting contextual materials that support video understandability over the long-term.
- Support meaningful future use of such videos by linking them with other related

VidArch Goals

- Develop a curator's workbench to assist in extending documentation strategies to easy collection of web-based video content.

Test Cases

- We are collecting videos related to the 2008 presidential elections as a focus but the goal is to have generally applicable tools.
- Several other areas under exploration as well including energy, disasters, and epidemics.

Archival Appraisal

- Appraisal is the area of the greatest professional challenge to the archivist.
  - Archivists' first responsibility is the selection of records of enduring value. – SAA, GAP Report, 1986.
  - The historical, cultural, and scientific record hinges on selection and appraisal.
  - Because of the unique nature of records, destruction is irrevocable.
  - Legal responsibilities.

Appraisal

- "In an archival context, appraisal[1] is the process of determining whether records and other materials have permanent (archival) value."
  - *SAA Glossary*
- "…any selection activity that enables archivists to identify recorded information that has enduring value, primarily for the documentation of modern society."
  - Cox & Samuels, 1988

Archival Documentation Strategies

- A methodology that guides selection and assures retention of adequate information about a specific geographic area, a topic, a process, or an event that has been dispersed throughout society. –*SAA Glossary*

Origins of Documentation Strategies

- "What constitutes the documentary heritage of today and how does a documentary heritage come into being? … it suffices to define the documentary heritage … as the totality of the existing evidence of historical activity, or as all the surviving documentation on past events."
  - Hans Booms, "Society and the Formation of a Documentary Heritage: Issues in the Appraisal of Archival Sources," 1972. (Eng. Trans. In 1987)

Origins of Documentation Strategies

- “A modem, complex, information-rich society requires that archivists reexamine their role as selectors… Archivists are challenged to select a lasting record, but they lack techniques to support this decision-making. Documentation strategies are proposed to respond to these problems.”
  - Samuels, “Who Controls the Past,” 1986.

Potential YouTube Impact

YouTube Basics

- Founded in February 2005.
- Allows people to easily upload and share video clips on www.YouTube.com and across the Internet through websites, mobile devices, blogs, and e-mail.
- Bought by Google in November 2006.
- WSJ, 8/30/06, reported:
  - > 6,000,000 videos.
  - > 45 TB of storage
  - > 1,800,000,000 views
  - Time spent watching > 9,000 years

YouTube Spread

- Time taken to spread to 50 million users:
  - Radio 17 years
  - TV 13 years
  - Internet 5 years
  - MySpace 3 years
  - YouTube 1 year
    - Jose-Marie Griffiths, September 2007

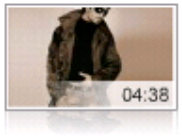
YouTube
Broadcast Yourself™

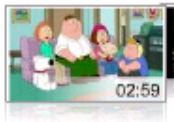
Sign Up | My Account | History | Help | Log In | Site:

Videos | Categories | Channels | Community | Upload

Search | powered by Google

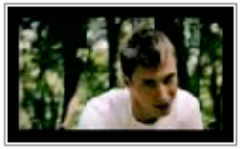
Videos being watched right now...





Promoted Videos

Prom Queen: Summer ...
promqueenty


DKA "Jak mam ufać s...
Grupa13

Japanese Robot Mont...
TokyoDV

Highlights: U.S. MN...
ussoccerdotcom

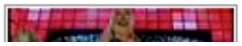
Bob Dylan "A Long Ride with the Right Music"
From: MyCadillacStory
Comments: 96

Login to rate video
225 ratings

Most Viewed Videos

See More Most Viewed Videos

Featured | **Most Viewed** | Most Discussed | Top Favorites


Britney Spears Gimme More -

From: xjvlx85

Login | Sign Up | Help

YouTube
Broadcast Yourself™

Sign Up | My Account | History | Help | Log In | Site:

Videos | Categories | Channels | Community | Upload

Search | powered by Google

Videos

Most Viewed : (Today) : All

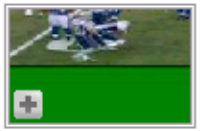
Browse
Most Recent
Most Viewed
Top Rated
Most Discussed
Top Favorites
Most Linked
Recently Featured
Most Responded
Watch on Mobile

Time
Today
This Week
This Month
All Time

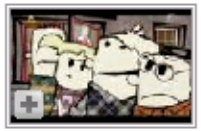
Category
All
Autos & Vehicles
Comedy

Apple iPhone Testing !
05:20
Added: 11 hours ago
From: iphonebest
Views: 318,644
71 ratings

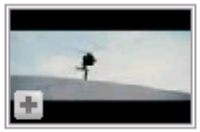
Barcelona 3 - 0 Betis. Resumen
02:56
Added: 1 day ago
From: jonaloide03
Views: 226,436
159 ratings



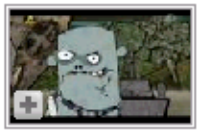
Indy Cheats
01:10
Added: 1 day ago
From: rrsafety
Views: 178,857
418 ratings



Wiatcy Móch - Ekskumancja cz.1
09:25
Added: 1 day ago
From: milanoss
Views: 165,416
96 ratings



Iron Man
00:59
Added: 1 day ago
From: JediMasterCaboose
Views: 132,703



Wiatcy Móch - Ekskumancja cz.2
06:06
Added: 1 day ago
From: milanoss
Views: 123,151

Very FUNNY condom commercial! LOL!!
01:45
Added: 22 hours ago
From: mycoolclipz
Views: 123,381

strutzi
01:45
Added: 1 day ago
From: onekidforall
Views: 119,869

SEE TOP SEARCH RESULTS ON ASK.COM

Celebrities
Aishwarya Rai
Akon
Britney Spears
Miley Cyrus
Heidi Klum
Kanye West

Appraisal and the Web

- Archivists, whether they are documenting an institution or collecting topically to provide a record of social phenomenon or personal accomplishment, must develop appraisal and collecting frameworks that take web-based materials into account.

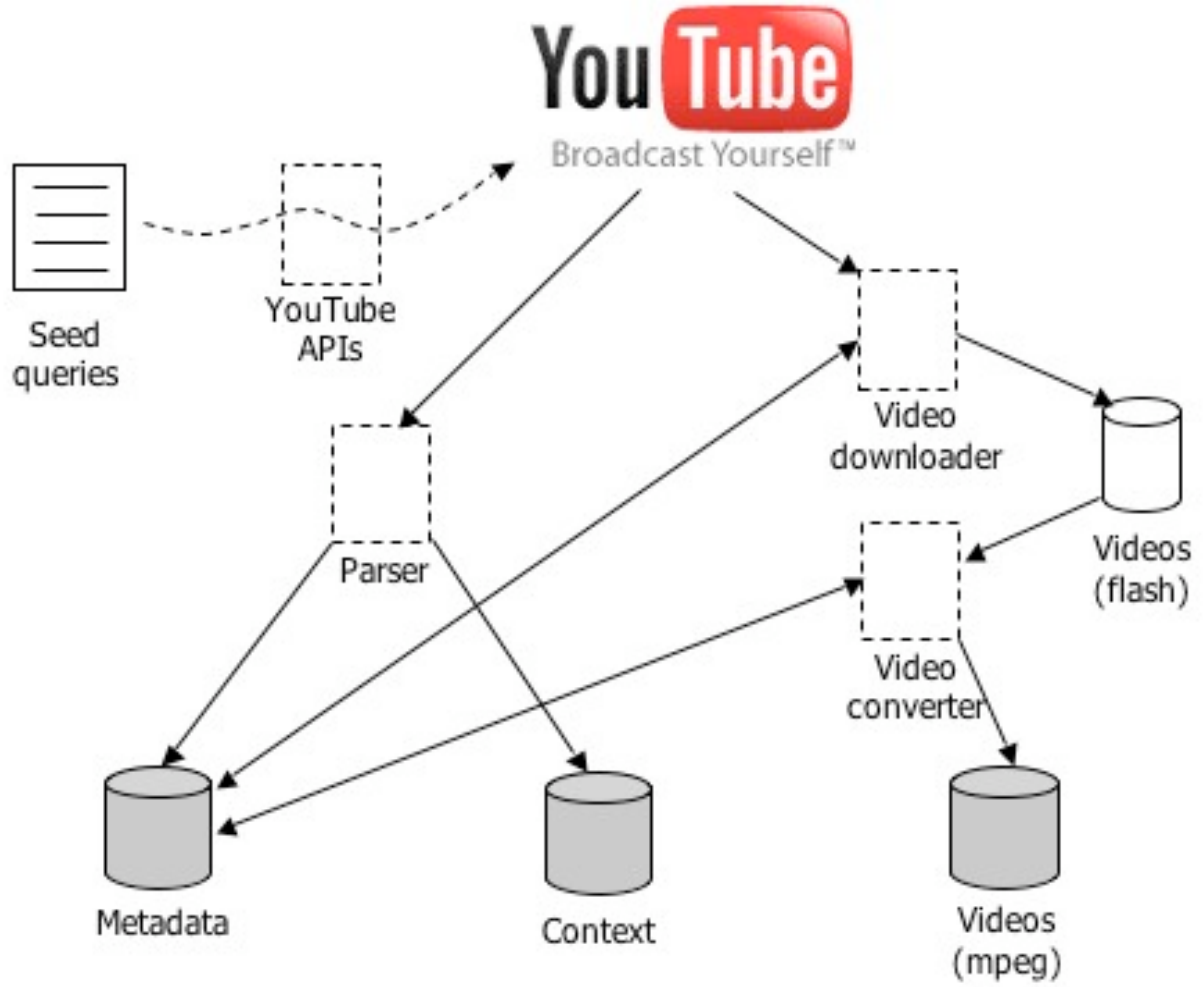
Opportunities of Extending Documentation Strategies to Web Videos

- Power of the masses to produce documentation of a subject or issue.
- Web-based materials can have a huge impact on society, including phenomenon such as voting behavior.
- Democratizes collecting strategies; more than network news or campaign materials being collected.
- Materials never before created or collected.

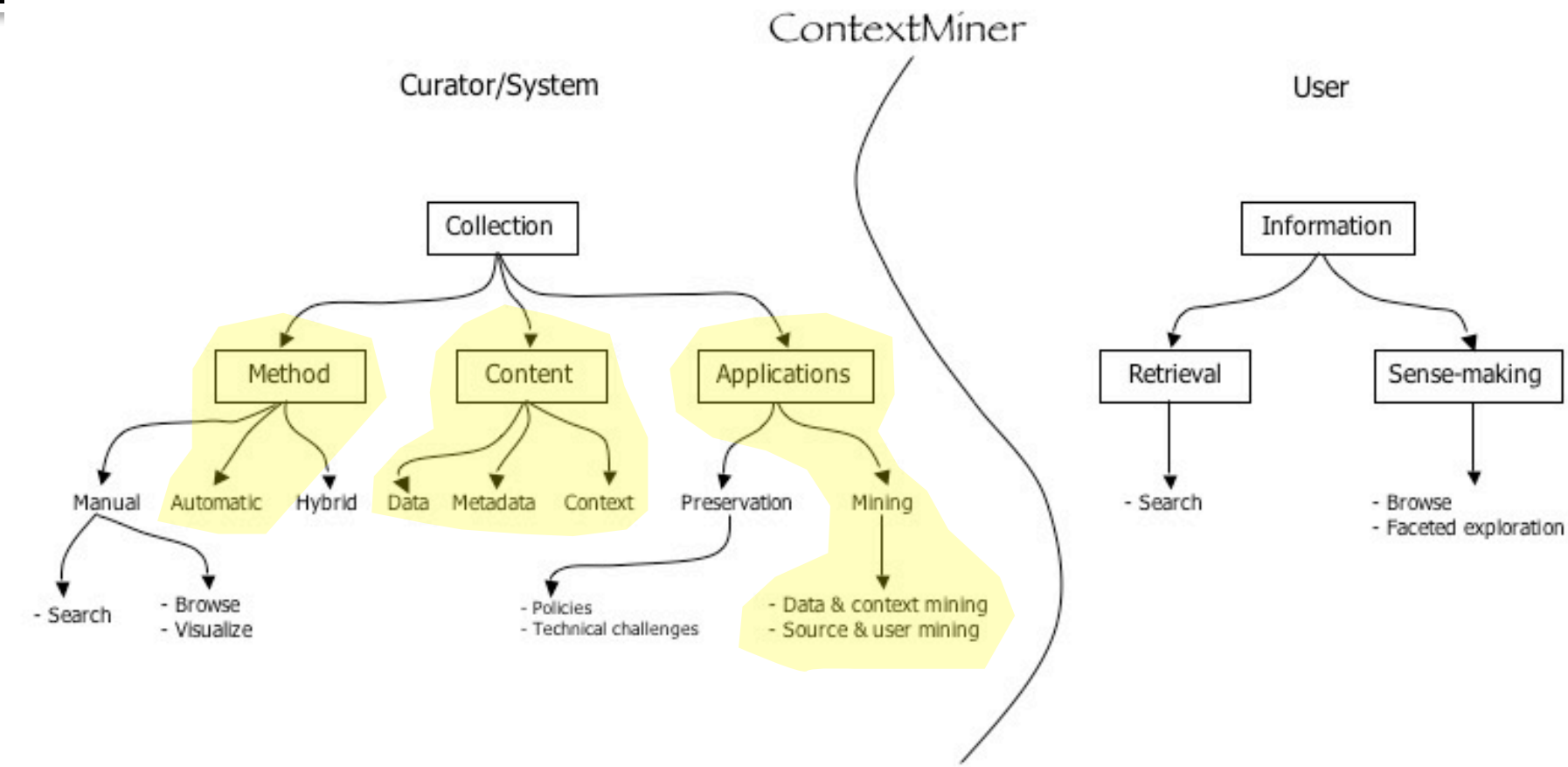
Challenges of Extending Documentation Strategies to Web Videos

- Potentially too many for hand-crafted, artisan approach.
- Ephemeral nature – here today; gone tomorrow.
- Variable quality and relevance.
- General lack of metadata.
- Unclear provenance and authenticity.
- Lack of contextualizing information.
- Just what is it we're collecting???

Video Harvesting from YouTube



ContextMiner - the big picture



Video Harvesting from YouTube for Election 2008

- 6 general queries and name searches
- Use You Tube APIs, screen scraping and other tools to collect videos
- Crawl everyday (almost)
- Get top 100 results for each query
- Collect more than 20 attributes (including all the comments)
- Download flash videos

Overview of the collection (as of 09/17/2007)

- Crawls = 100
- Unique videos > 10,000
- Video files > 100 GB
- Total honors > 200
- Total views > 125,000,000
- Total ratings > 800,000
- Total comments ~ 800,000

Election 2008 Collecting Scenario

- Curator of Hillary Clinton's campaign.
- Direct feed of materials from Clinton's staff and Democratic Party.
  - Press releases, video, interviews, Face Book, etc.
- Wide variety of traditional media – newspapers, TV, radio.
- Now wide variety of bottom-up materials, including You Tube videos
  - "Official" CNN debate videos, reactions, etc.

THIS IS THE INTERNET ADDRESS OF MY WEBCAM REACTION TO A YOUTUBE PARODY OF MY LATEST AD
ELECT
ME
ME
WASSERMAN
© 2007 THE BOSTON GLOBE
DIST. BY TRIBUNE MEDIA SERVICES

Next Steps

- Develop specific use cases for the Context Miner.
- Test cases with curators.
- Refine our tools based on this testing with regards to retrieval, mining of metadata, etc.
- Learn more about the YouTube and Google Video environments.
- Collect other information sources such as blogs that relate to the video content.

Thank you!

- Helen R. Tibbo
- School of Information and Library Science
- University of North Carolina at Chapel Hill
- tibbo@ils.unc.edu
- http://ils.unc.edu/vidarch