Archives Unleashed

Summer 2019 Newsletter

Summer is flying by! Don't miss out on this newsletter edition of project highlights and announcements.

The Archives Unleashed team has had a busy year and we are proud to share our work with you!

A complete list of all publications is available at our publication page.

Celebrating with our Collaborators

The Archives Unleashed Team would like to extend special thanks to two doctoral students: Sarah McTavish and Rebecca MacAlpine. They have done a phenomenal job of enhancing our Cloud Learning resources by creating the new Notebook tutorial and working on a series of pedagogical materials for running web archive analysis workshops.

Special Announcements

Full details on work towards the 0.18.0 release available:
https://github.com/archivesunleashed/aut/compare/aut-0.17.0...master

CEDWARC

Our colleagues at the Continuing Education to Advance Web Archiving Project (CEDWARC) are offering a free in-person training workshop this fall in Washington DC.

Call for participation: CEDWARC Workshop & Travel Grants

Supported by the IMLS LB21 program, the Continuing Education to Advance Web Archiving (CEDWARC) project develops a continuing education curriculum and teaches library and archive professionals advanced web archiving and analysis techniques.

The workshop offers advanced training to working library and archive professionals. Subject to availability, students in the related fields are also welcome. The training modules are customized to the knowledge and skill levels of the intended audience and can be integrated into their job responsibilities, as well as prepare them for future library and archive service developments.

Built on the Web Archiving Fundamentals module, participants will be taken on a problem-solving tour through five additional modules, including Storytelling, Social Feed Manager, ArchiveSpark, Archives Unleashed, and Events Archiving. Facilitators of the workshop include web archiving experts and tool developers from Virginia Tech, Los Alamos National Lab, Old Dominion University, George Washington University, the Internet Archive, the University of Waterloo, and Rhizome.

About 15 travel grants are also available to assist participants' travel expenses.

Date: Oct 28, 2019
Location: Gelman Library, George Washington University, 2130 H St NW, Washington, DC 20052
Cost: Free
Travel Grant Application Deadline: Aug 30, 2019

For details on workshop registration and travel grant application, please visit the project website at: http://cedwarc.github.io/

Please note, travel grants are available to US residents only

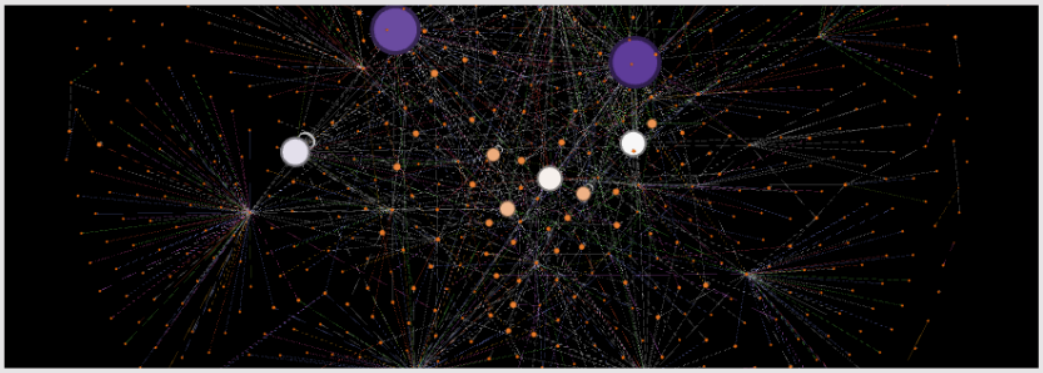
We will be hosting our final datathon in the Spring of 2020. Watch out for the Call for Participation opening this fall.

What's been happening?

- Toolkit development
  - Working towards 0.18.0 release
  - Dependency work *(reducing code dependencies in the Toolkit, we continue to work towards long-term sustainability of the code base)*
  - Strengthening our data frame implementation
  - Closed 37+ issues
- Cloud development and testing
- Launched the Archives Unleashed Cloud Notebook
- Developed a new learning guide for working with Archives Unleashed Notebooks (by Sarah McTavish)
- Created and launched Archives Unleashed Project Video
- Presented at a number of conferences and scholarly meetings
- Continued work on sustainability planning
- Hosted our third datathon in Washington, DC
- Initiated final datathon planning, stay tuned for CFP in early fall

The Archives Unleashed Cloud turned 1 this past April.
Check out these usage stats! *(updated August 2019)*

ARCHIVES UNLEASHED CLOUD



Helping researchers and scholars access and analyze petabytes of historical internet content.

EST. 2018

221 TB
Data Analyzed

983
Collections

5126
Jobs Completed

178
Users

Featured Articles

Perspectives: Navigating Web Archives

High school student experience highlights the opportunities for the web archiving community to engage long scholars.

(Read More...)

On the Road with the Archives Unleashed Team

The Archives team has been on the road, meeting up with colleagues during conference season.

(Read More...)

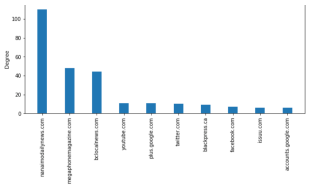
Check It Out

Archives Unleashed Cloud Learning Guides

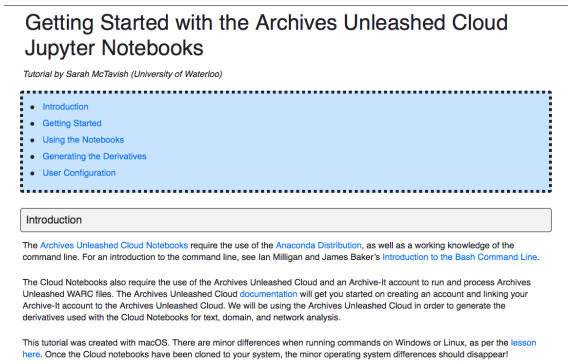
We are pleased to introduce a new learning guide, Getting Started with the Archives Unleashed Cloud Jupyter Notebooks. This guide provides a basic walkthrough for integrating data from the Archives Unleashed Cloud with our Cloud Notebooks.

The example dataset provided was created with the Archives Unleashed Cloud, and is drawn from the B.C. Teachers' Labour Dispute (2014), collected by the University of Victoria Libraries.

You can take Archives Unleashed Notebooks for a spin: https://github.com/archivesunleashed/auk-notebooks



In [16]:
```
plt.rcParams['figure.figsize'] = [25, 25] # set the figure size for the graph

# Convert rgb values to between 0 & 1 and put them into a numpy array.
rgbs = zip([x[1]/255 for x in graph.nodes('r')], [x[1]/255 for x in graph.nodes('g')], [x(
colormap = [np.array(x) for x in rgbs]

# Labels
mapping = {x[0]: x[1] for x in graph.nodes('label')}

# Use Archive Unleashed Clouds Positions (saves on load time)
zippos = zip(graph.nodes, [x[1] for x in graph.nodes('x')], [x[1] for x in graph.nodes('y'
positions = {x[0]: np.array([x[1],x[2]]) for x in zippos}

# Node sizes based on degree
size = np.array([x[1] * 100 for x in graph.nodes('size')])

# Draw the graph
nx.draw(graph, pos=positions, show_labels=True, labels=mapping, font_size=10, node_size=si
plt.show()
```



Getting Started with the Archives Unleashed Cloud Jupyter Notebooks

Tutorial by Sarah McTavish (University of Waterloo)

- Introduction
- Getting Started
- Using the Notebooks
- Generating the Derivatives
- User Configuration

Introduction

The Archives Unleashed Cloud Notebooks require the use of the Anaconda Distribution, as well as a working knowledge of the command line. For an introduction to the command line, see Ian Milligan and James Baker's Introduction to the Bash Command Line.

The Cloud Notebooks also require the use of the Archives Unleashed Cloud and an Archive-It account to run and process Archives Unleashed WARC files. The Archives Unleashed Cloud documentation will get you started on creating an account and linking your Archive-It account to the Archives Unleashed Cloud. We will be using the Archives Unleashed Cloud in order to generate the derivatives used with the Cloud Notebooks for text, domain, and network analysis.

This tutorial was created with macOS. There are minor differences when running commands on Windows or Linux, as per the lesson here. Once the Cloud notebooks have been cloned to your system, the minor operating system differences should disappear!

Introducing the Archives Unleashed Project

We have a new video! Learn about our project's beginnings and the different ways you can use the Archives Unleashed Toolkit.

Get Involved

Interested in getting involved with the Archives Unleashed Project? Connect with our team and help grow our community

- Follow us on Twitter
- Join our Slack group
- Participate on Github
- Subscribe to our newsletter
- Connect and tell us how you've used our tools and platforms
- Share our news with colleagues and friends

Twitter GitHub Website Email

The Archives Unleashed Project aims to make petabytes of historical internet content accessible to scholars and others interested in researching the recent past. Supported by a grant from the Andrew W. Mellon Foundation, the project will develop web archive search and data analysis tools to enable scholars and librarians to access, share, and investigate recent history since the early days of the World Wide Web.

The Archives Unleashed Project
200 University Avenue West I Waterloo, ON N2L 3G1

Copyright © 2018 The Archives Unleashed Project. All Rights Reserved.

To change how you receive these, you can update your preferences or unsubscribe from this list.

This email was sent to << Test Email Address >>

why did I get this? unsubscribe from this list update subscription preferences
Archives Unleashed · 200 University Ave W · Waterloo, On N2L 3G1 · Canada