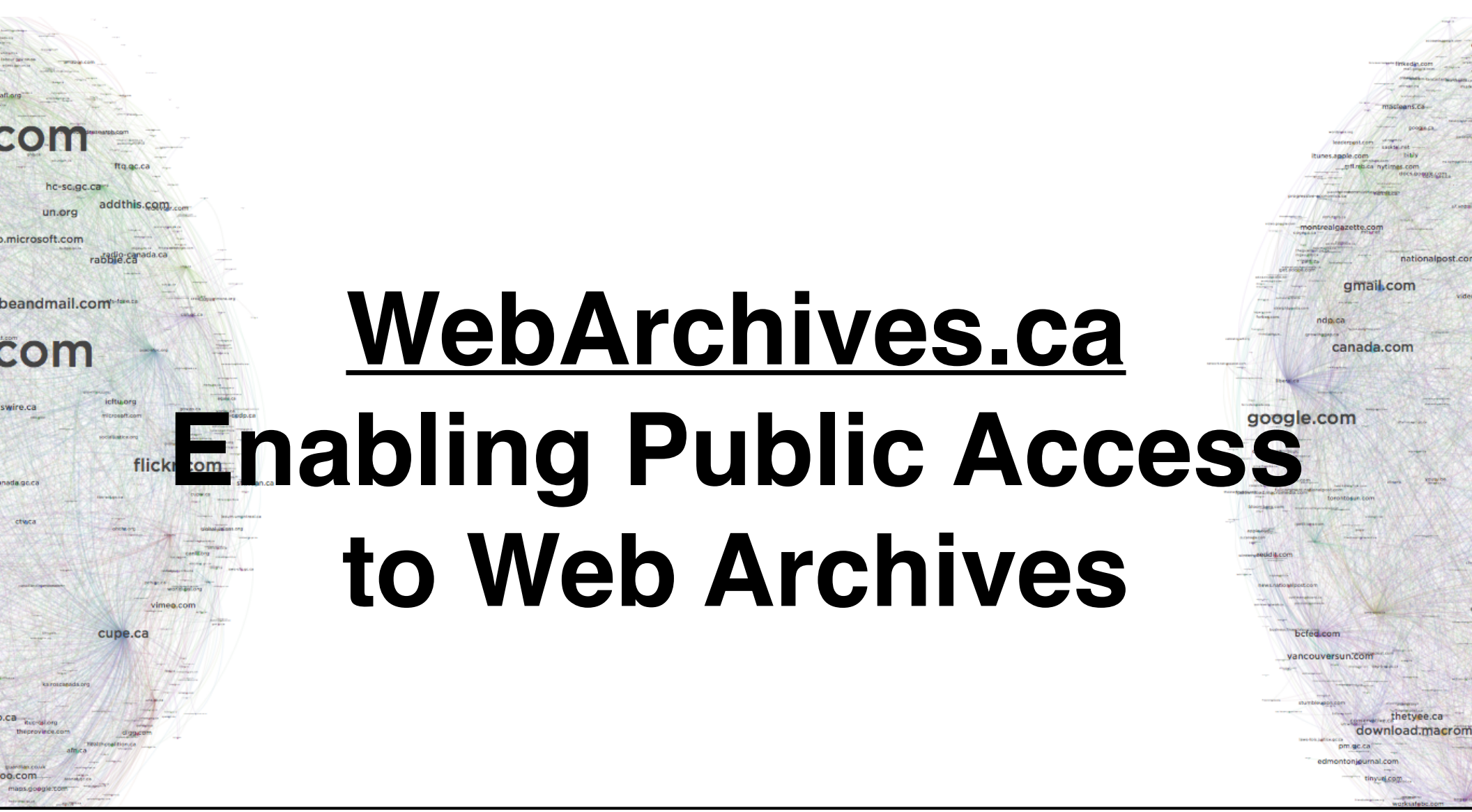
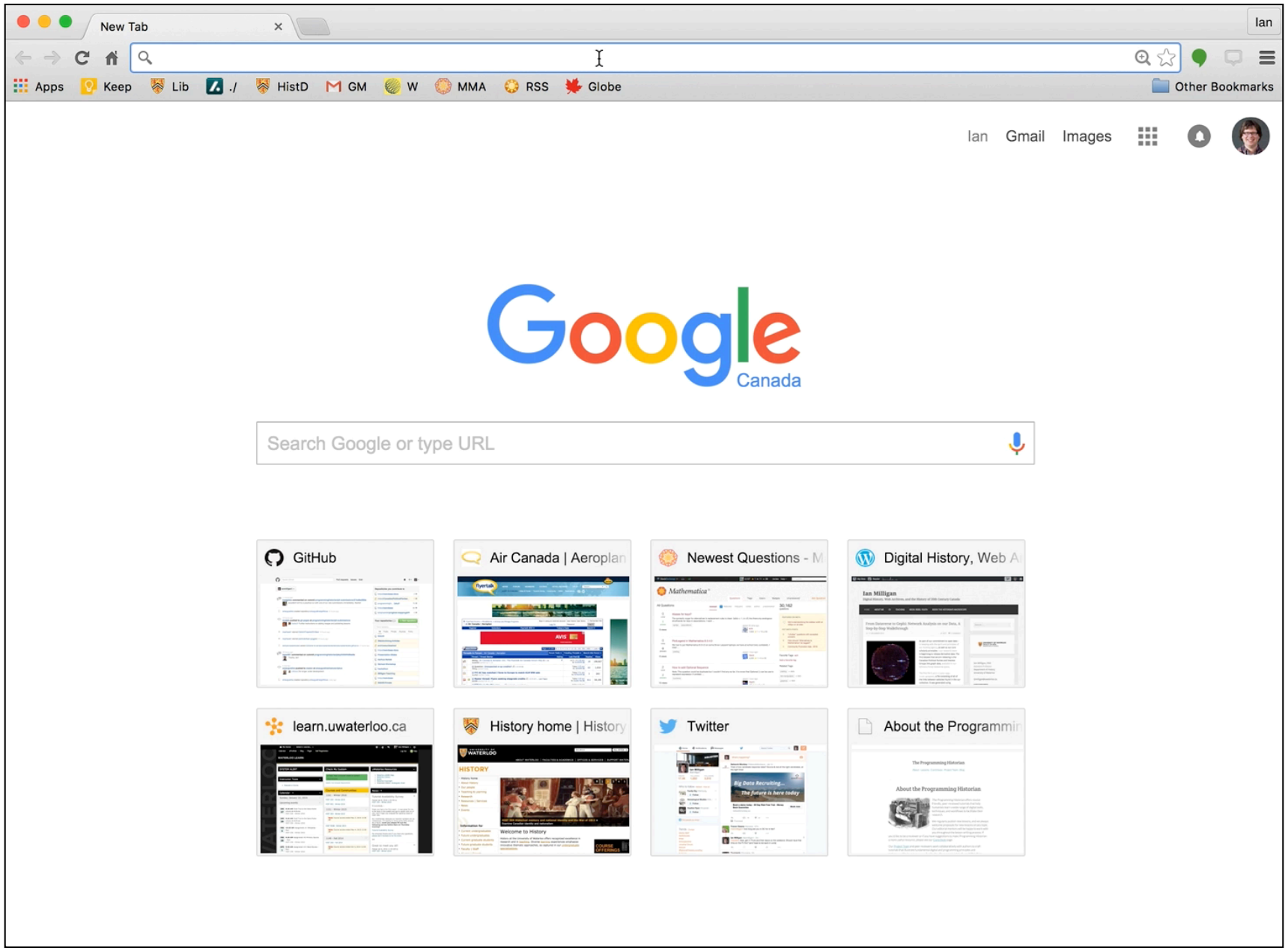


WebArchives.ca
Enabling Public Access to Web Archives

Ian Milligan
Assistant Professor, History

UNIVERSITY OF WATERLOO

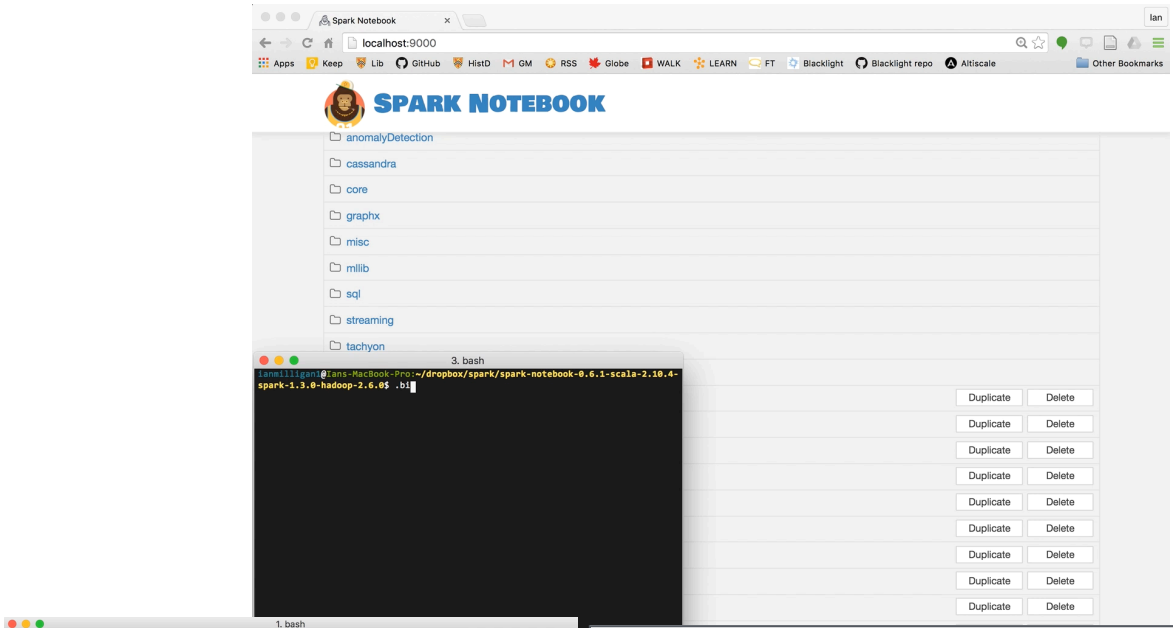
Web Archives are great – but how do you use them?

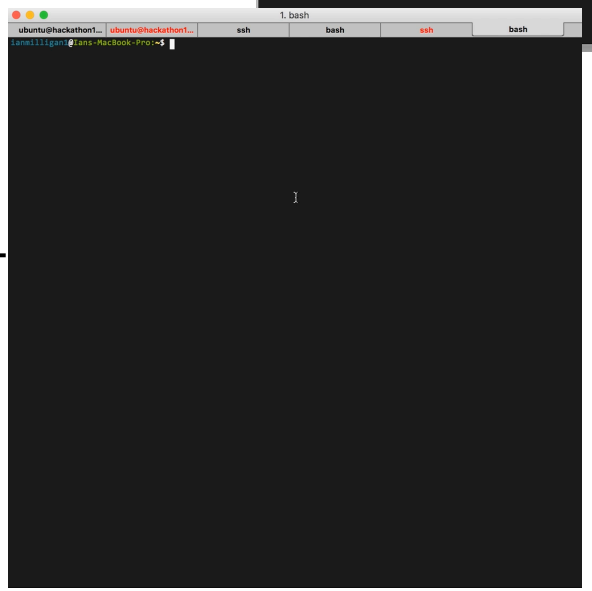


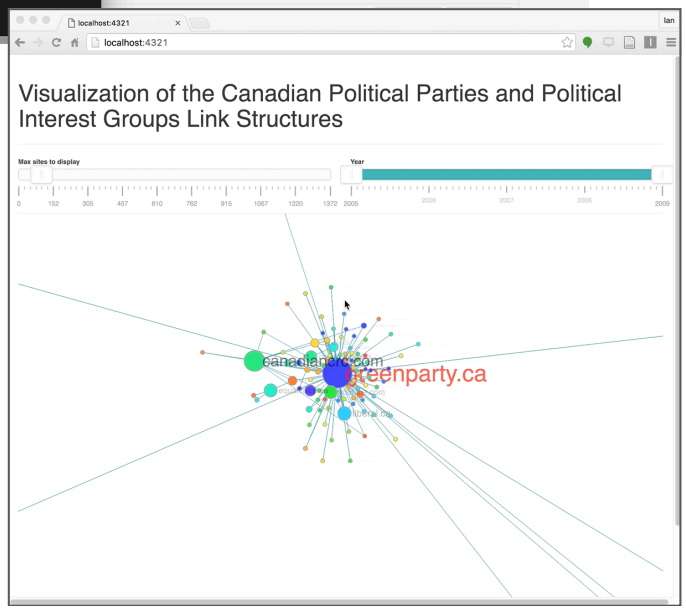
Limited to close, slow reading.

Two Models

- **The coding, computational approach discussed in the last session (warcbase)**
- warcbase.org
- docs.warcbase.org
- If you have the files storage, etc.







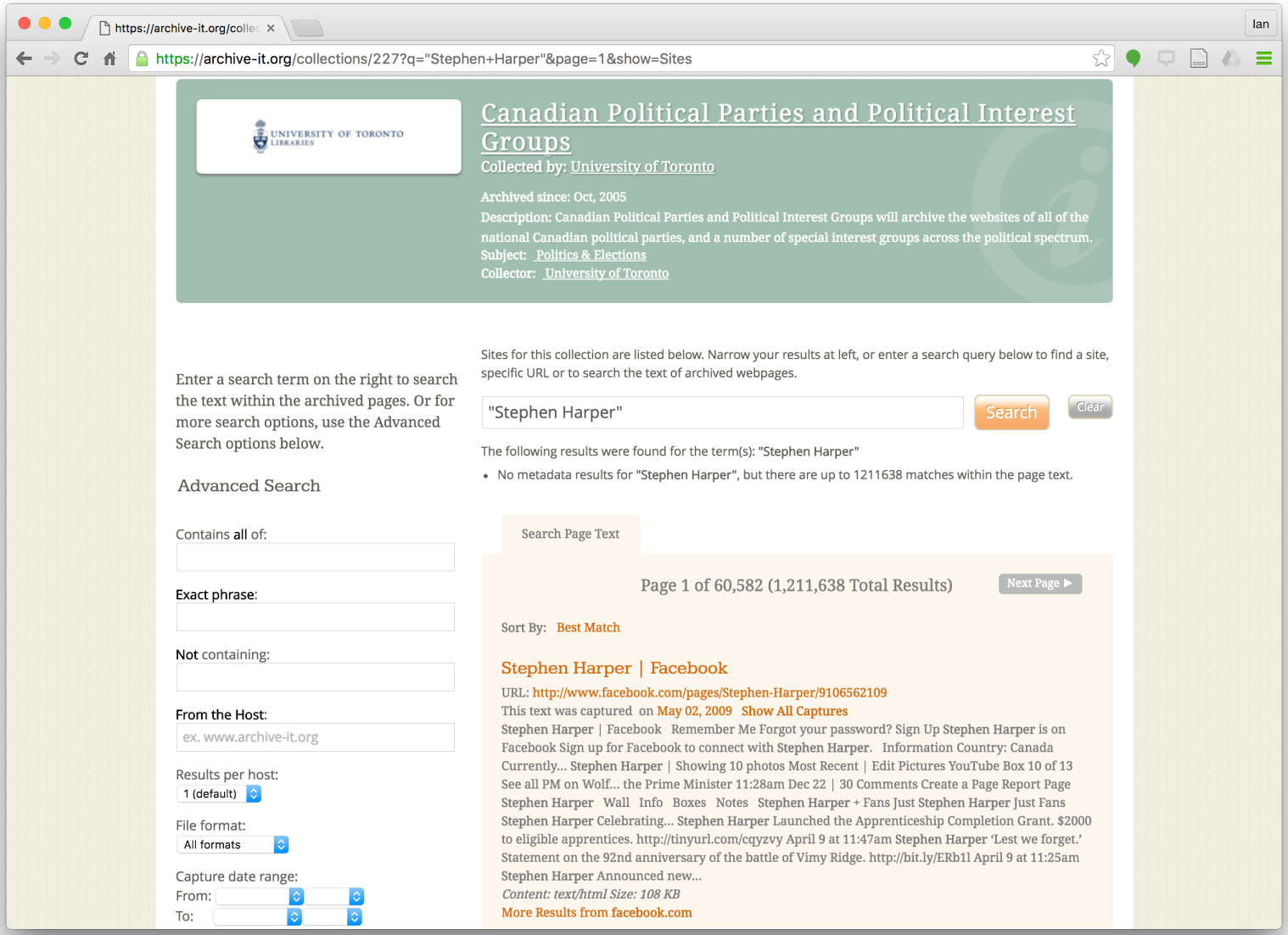
Lighter-weight interfaces needed for public, occasional research use.

Andrew Jackson, Jimmy Lin, Ian Milligan, and Nick Ruest, "**Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities**," *Joint Conference on Digital Libraries 2016*, Newark NJ.

In short: we will need search engines.

… but what will our search engines look like?

Standard SERP Inadequate



Overview first, zoom and filter, then details-on-demand.

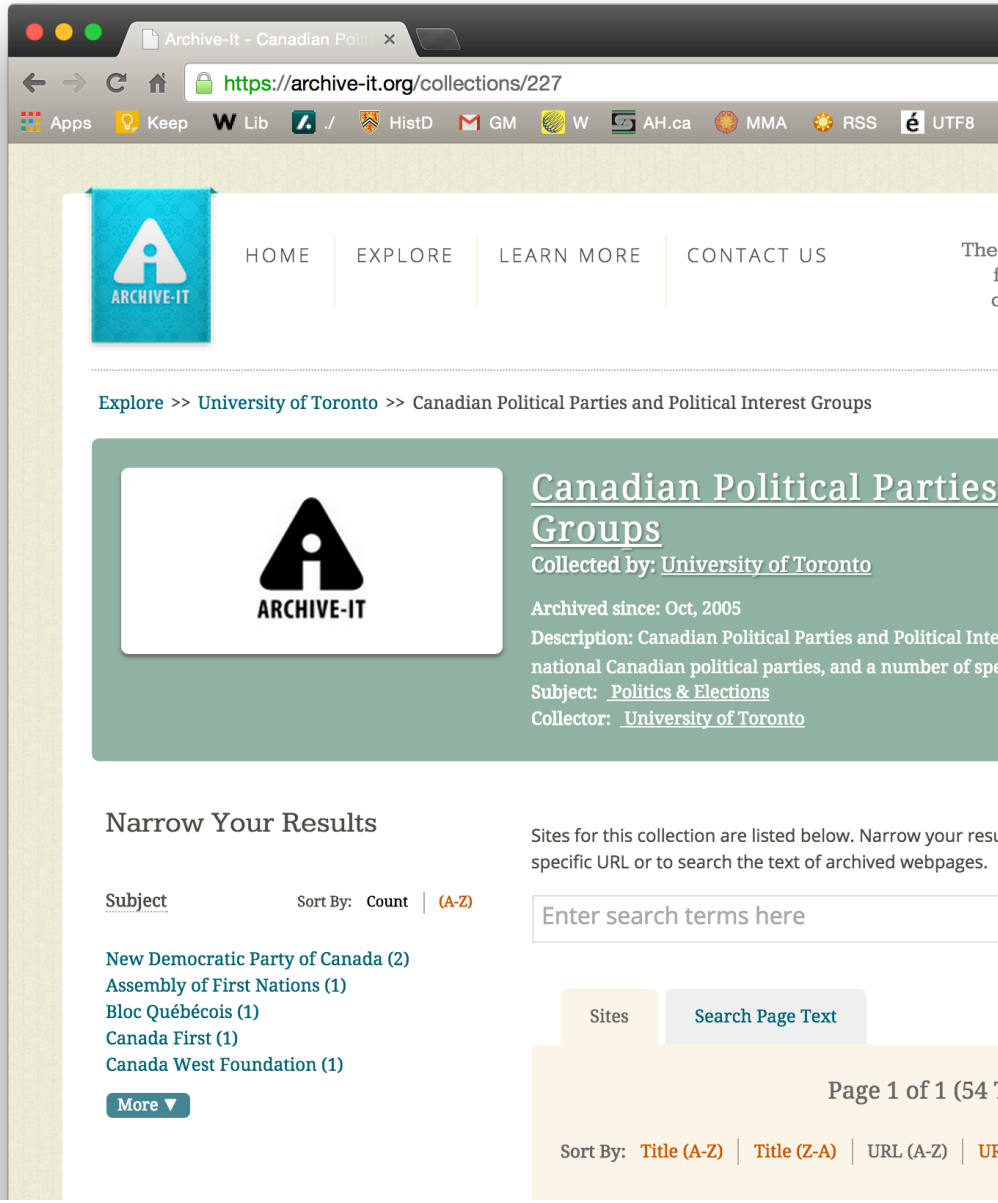
Schneiderman's mantra (1996)

Close - Medium - Distant

- **Distant Reading**: Billions of documents
- **Close Reading**: One document
- **“Middle game”:** Moving between these levels

Building Portals

- Democratizing access so that historians can use them.
- Building **transparent indexes**.
- But they have to be useful and tested…



Pivotal Changes in Canadian Politics, 2005-2015

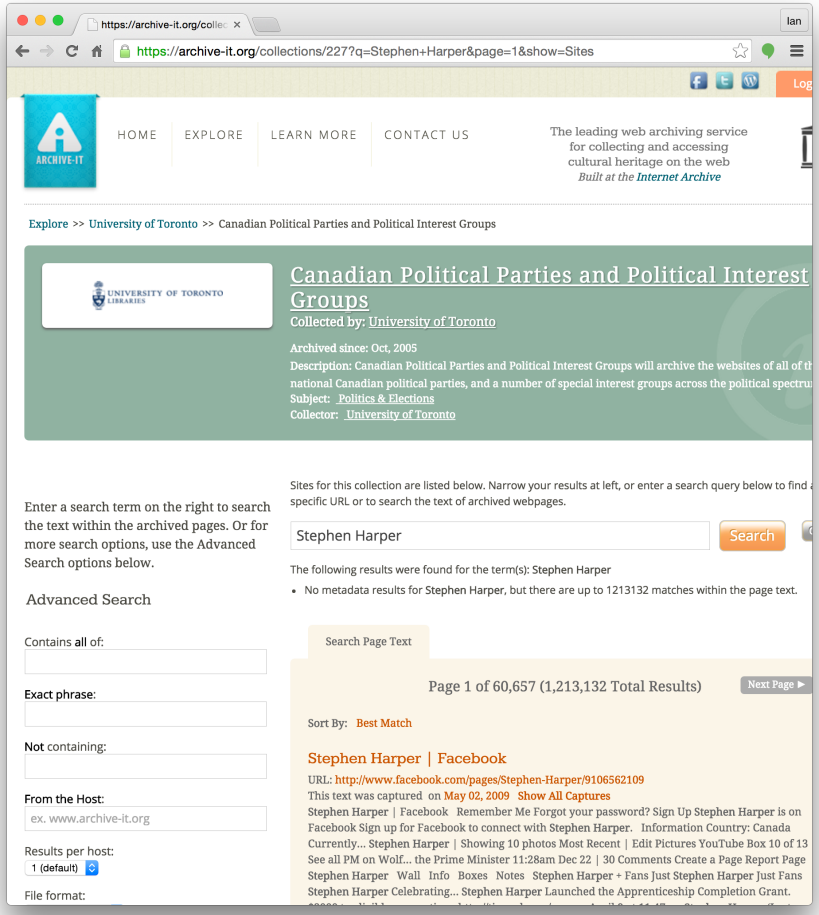
- Militarization of Canadian society?
- Change from ‘natural governing party’ of Liberals to Conservatives
- Major policy changes on foreign policy, environment, etc.
- Sweeping change to how we understand our history
- How to measure?

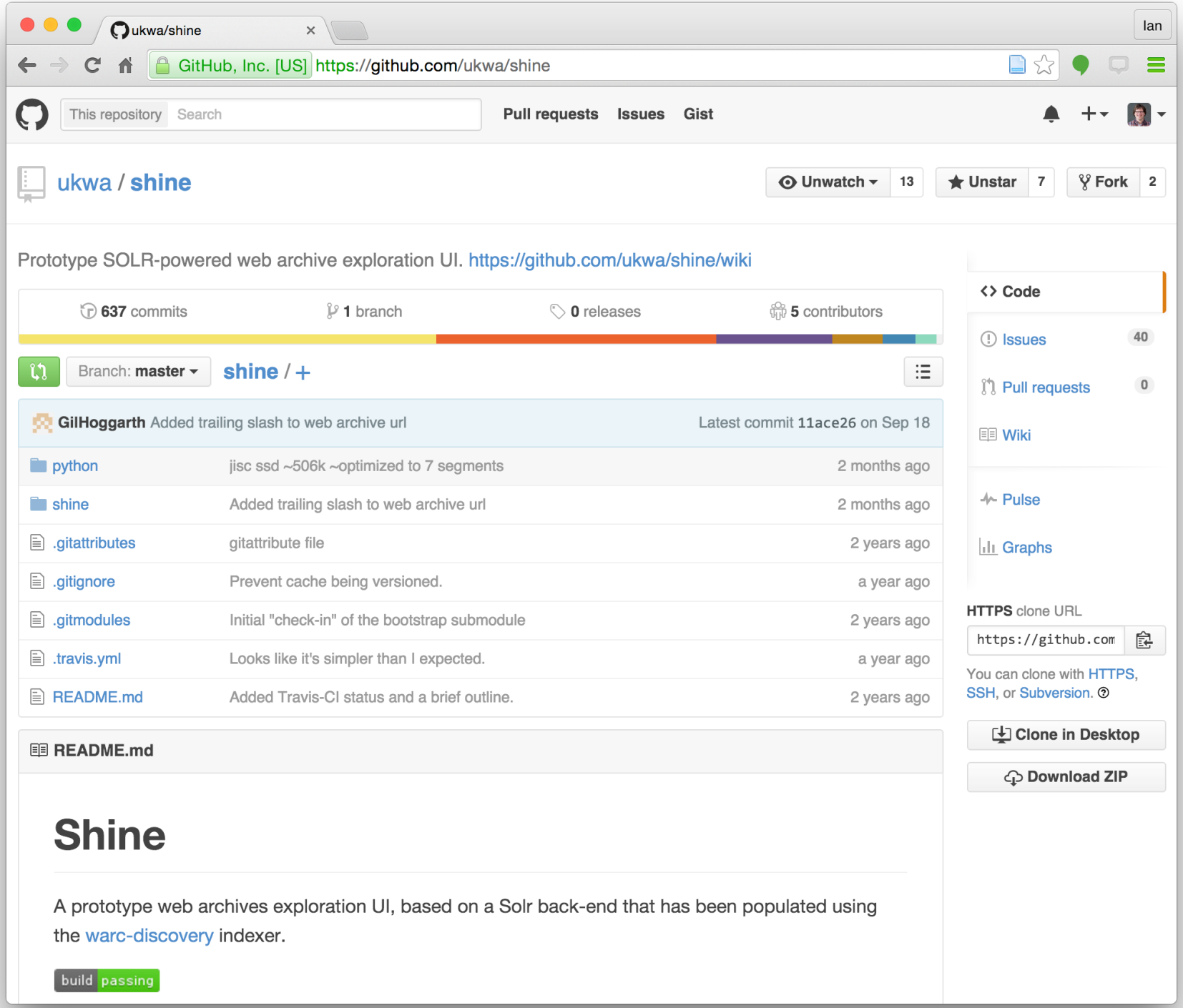
Canadian Political Parties & Political Interest Group Collection

- 50 Websites
  - All major political parties
  - Minor political parties
  - Political interest groups
- Collected quarterly between 2005 & present.

Current Interface

- Very limited - simple search engine, some advanced options; no facets
- Great collections.. but nobody uses them!




ukwa/shine
Ian
GitHub, Inc. [US] https://github.com/ukwa/shine
This repository Search
Pull requests Issues Gist
ukwa / shine
Unwatch 13 Unstar 7 Fork 2
Prototype SOLR-powered web archive exploration UI. https://github.com/ukwa/shine/wiki
637 commits 1 branch 0 releases 5 contributors
Branch: master shine / +
GilHoggarth Added trailing slash to web archive url
Latest commit 11ace26 on Sep 18
python jisc ssd ~506k ~optimized to 7 segments 2 months ago
shine Added trailing slash to web archive url 2 months ago
.gitattributes gitattribute file 2 years ago
.gitignore Prevent cache being versioned. a year ago
.gitmodules Initial "check-in" of the bootstrap submodule 2 years ago
.travis.yml Looks like it's simpler than I expected. a year ago
README.md Added Travis-CI status and a brief outline. 2 years ago
README.md
Shine
A prototype web archives exploration UI, based on a Solr back-end that has been populated using the warc-discovery indexer.
build passing
Code
Issues 40
Pull requests 0
Wiki
Pulse
Graphs
HTTPS clone URL
https://github.com
You can clone with HTTPS, SSH, or Subversion.
Clone in Desktop
Download ZIP

Great research question that our contemporary historians were studying (Canada changing)

+

Great collection (all the political parties + many interest groups)

+

Ope Source Software

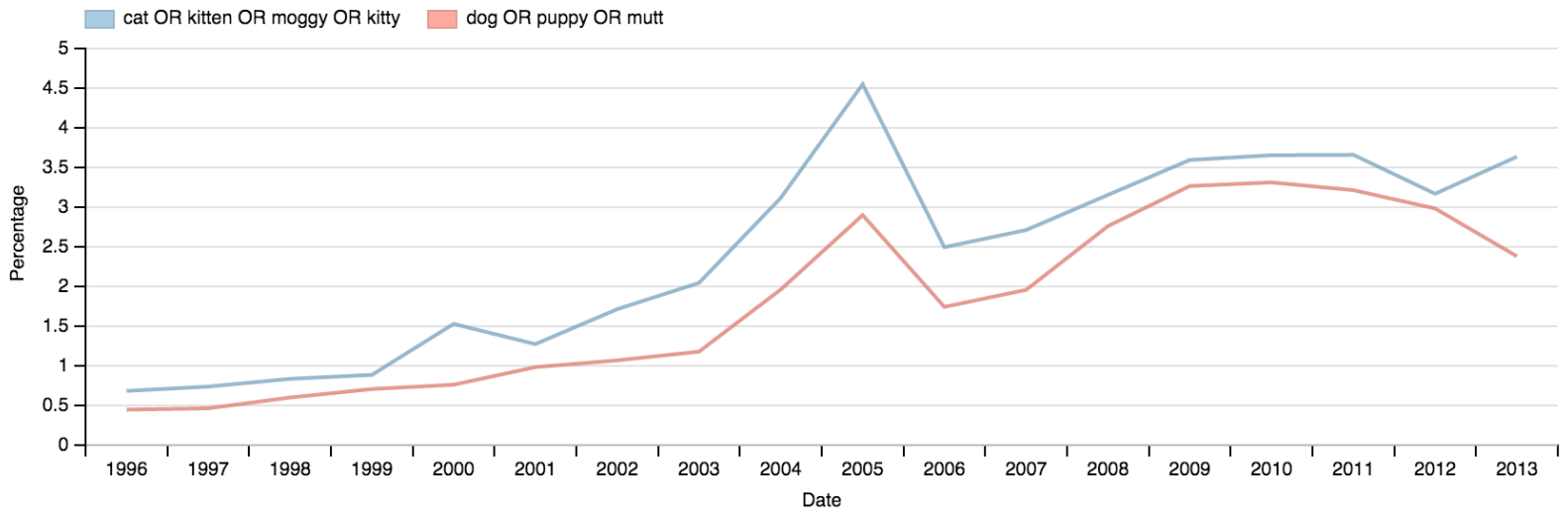
Trend results 1996-2013 fo

https://www.webarchive.org.uk/shine/graph?query=cat+OR+kitten+OR+moggy+OR+kitty%2C+dog+OR+pu...

UK Web Archive | Search | Trends | Login

Warning! This is a research prototype for a web archive search service, and may be taken down at any time.

cat OR kitten OR moggy OR kitty, dog OR | 1996 | 2013 | Update Graph | Reset All



Click on a point on the graph to show a random sample of the matching records from that year...

With Nick Ruest (York University), Nich Worby (University of Toronto), Jimmy Lin (University of Waterloo)

Welcome to the Web Archives for Historical Research political parties portal. Before diving in, we encourage you to visit our about page. ×

The Canadian Political Parties and Political Interest Groups Portal

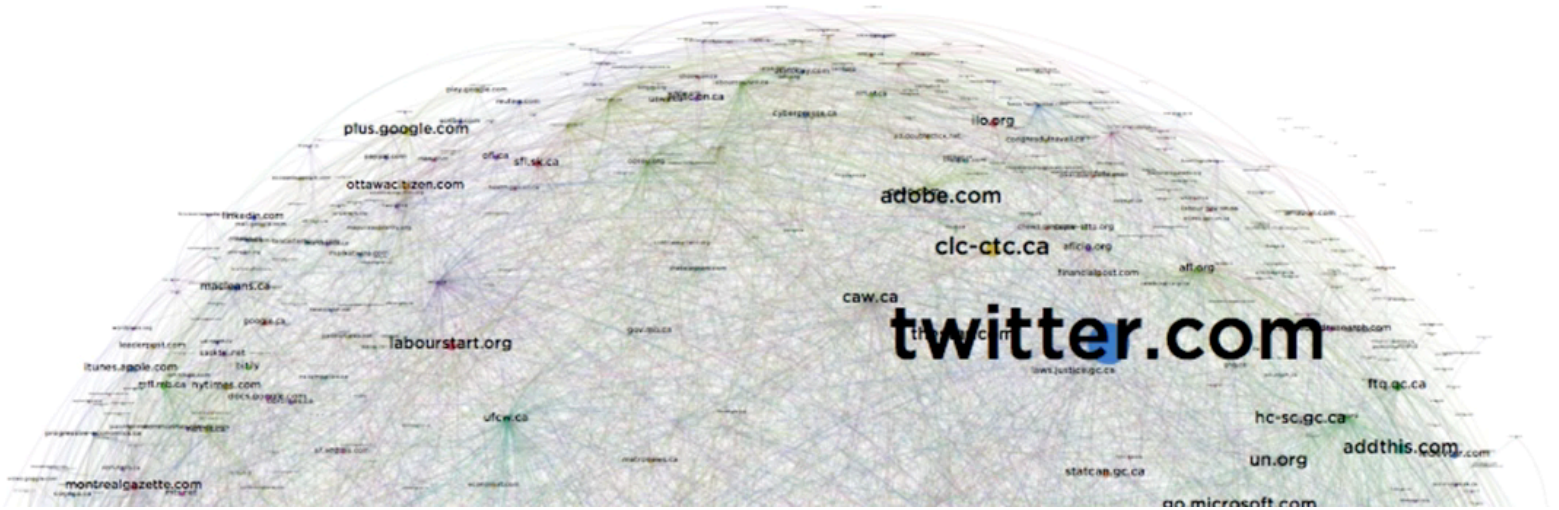
On this website, you can search web archived content from 50 political parties and political interest groups, from October 2005 to March 2015.

Curious how the Liberal Party of Canada responded to the 2008 financial crisis (a search for "recession" in 2008, liberal.ca)? How the Canadian Centre for Policy Alternatives reacted to Michael Ignatieff? Now you can check it all out.

Options include:

- **Basic keyword searching** [Example: "Rob Ford", only Liberal.ca]
- **Graphing trends over time** [Example: Liberal Opposition Leaders, 2005-2015]
- **Advanced search, including words in proximity to each other** [Example: environmental and tax within 25 words of each other]

Below, here are all of the links for the entire time period, visualized below.

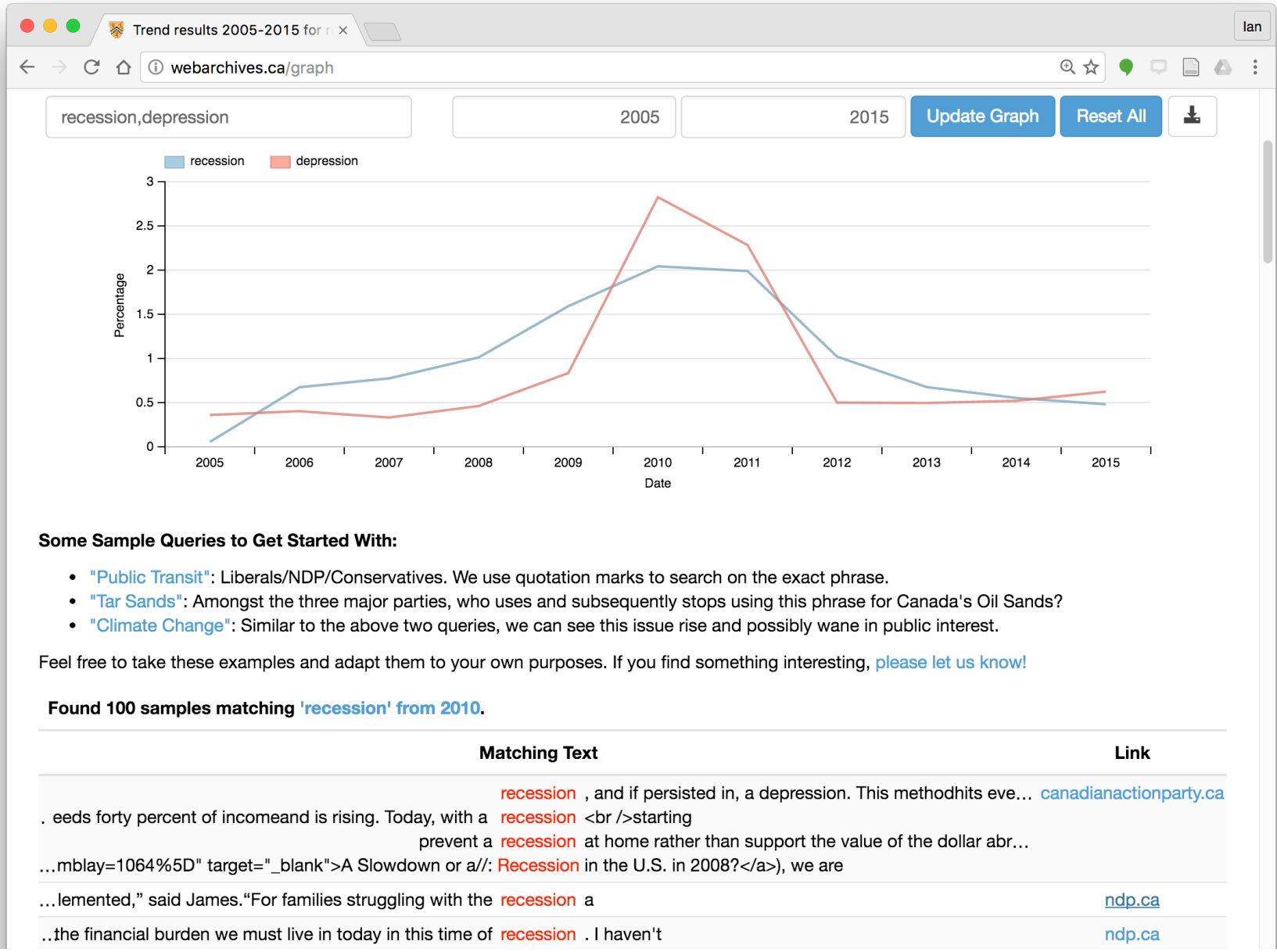


Trend Diagram

Distant =

Middle Game

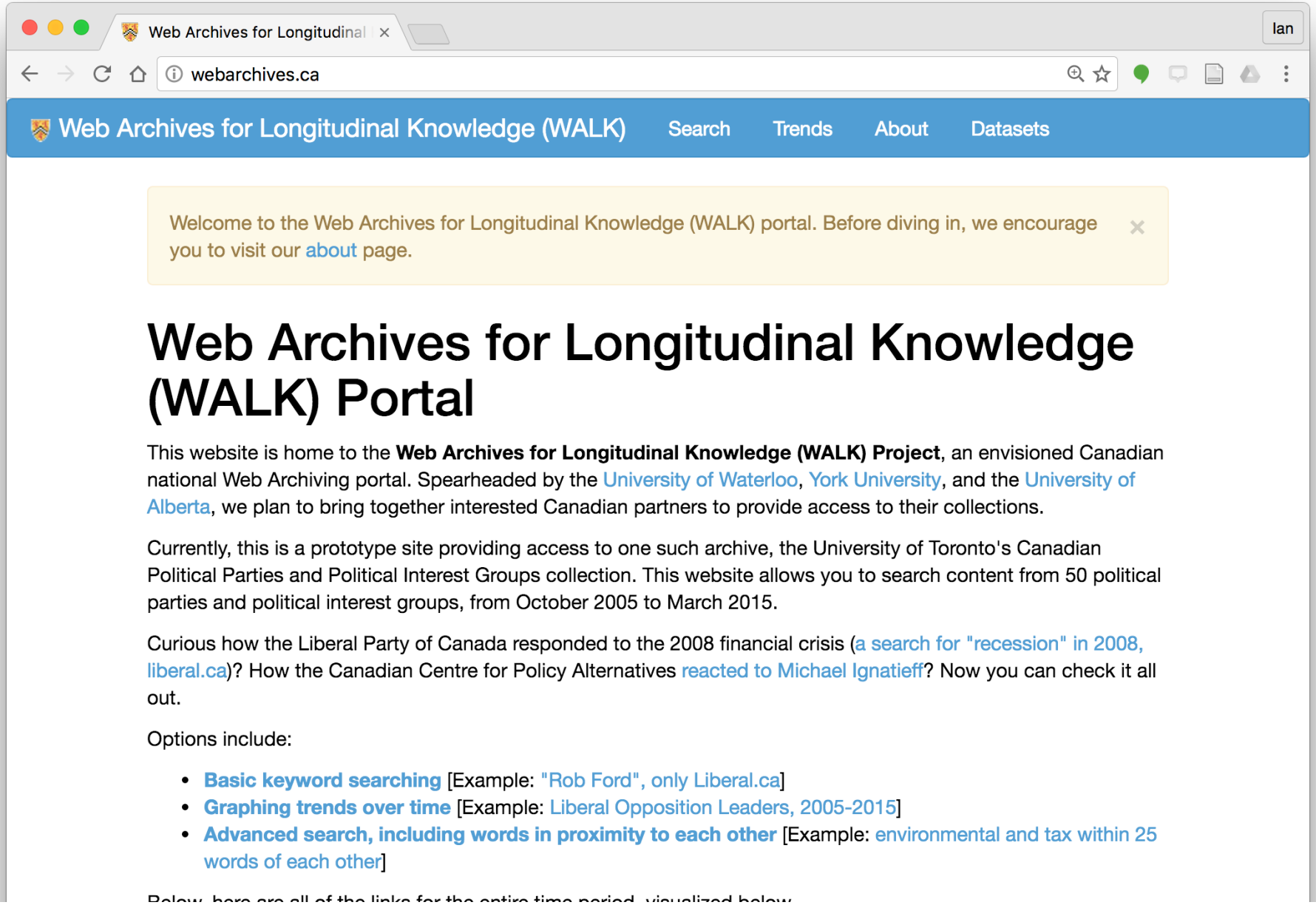
Close =



Five Things We Learned

- Political parties delete content
- User-generated comments were more common in political parties
- Absences can be more informative than presences
- We can see the rise/fall of prominent people
- Enabling user access is truly transformative

Next Steps


Web Archives for Longitudinal
webarchives.ca
Web Archives for Longitudinal Knowledge (WALK)
Search
Trends
About
Datasets
Welcome to the Web Archives for Longitudinal Knowledge (WALK) portal. Before diving in, we encourage you to visit our about page.
Web Archives for Longitudinal Knowledge (WALK) Portal
This website is home to the Web Archives for Longitudinal Knowledge (WALK) Project, an envisioned Canadian national Web Archiving portal. Spearheaded by the University of Waterloo, York University, and the University of Alberta, we plan to bring together interested Canadian partners to provide access to their collections.
Currently, this is a prototype site providing access to one such archive, the University of Toronto's Canadian Political Parties and Political Interest Groups collection. This website allows you to search content from 50 political parties and political interest groups, from October 2005 to March 2015.
Curious how the Liberal Party of Canada responded to the 2008 financial crisis (a search for "recession" in 2008, liberal.ca)? How the Canadian Centre for Policy Alternatives reacted to Michael Ignatieff? Now you can check it all out.
Options include:
Basic keyword searching [Example: "Rob Ford", only Liberal.ca]
Graphing trends over time [Example: Liberal Opposition Leaders, 2005-2015]
Advanced search, including words in proximity to each other [Example: environmental and tax within 25 words of each other]

Thanks!

compute canada | calcul canada

Ontario

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Ian Milligan

Assistant Professor, History

UNIVERSITY OF WATERLOO