Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities

---

Andrew Jackson
Web Archiving Technical Lead
@anjacks0n

Jimmy Lin
Professor
@lintool

Ian Milligan
Assistant Professor
@ianmilligan1

Nick Ruest
Digital Assets Librarian
@ruebot

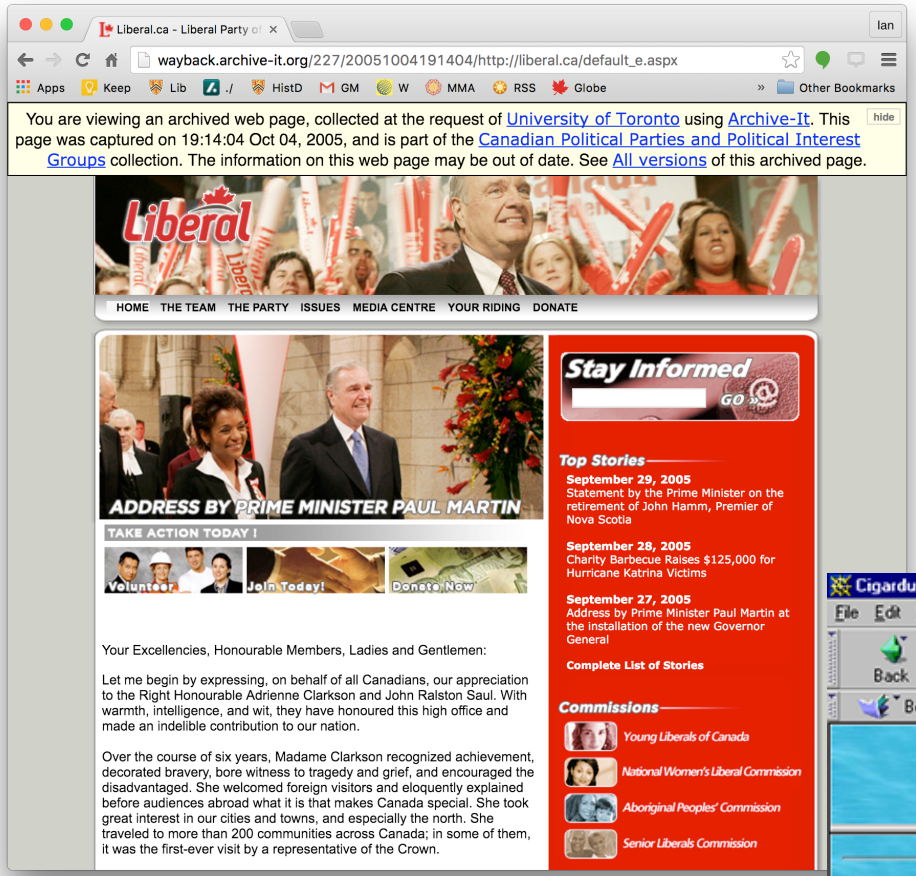
British Library

UNIVERSITY OF WATERLOO
FACULTY OF MATHEMATICS
David R. Cheriton School
of Computer Science

UNIVERSITY OF WATERLOO
FACULTY OF ARTS
Department of History

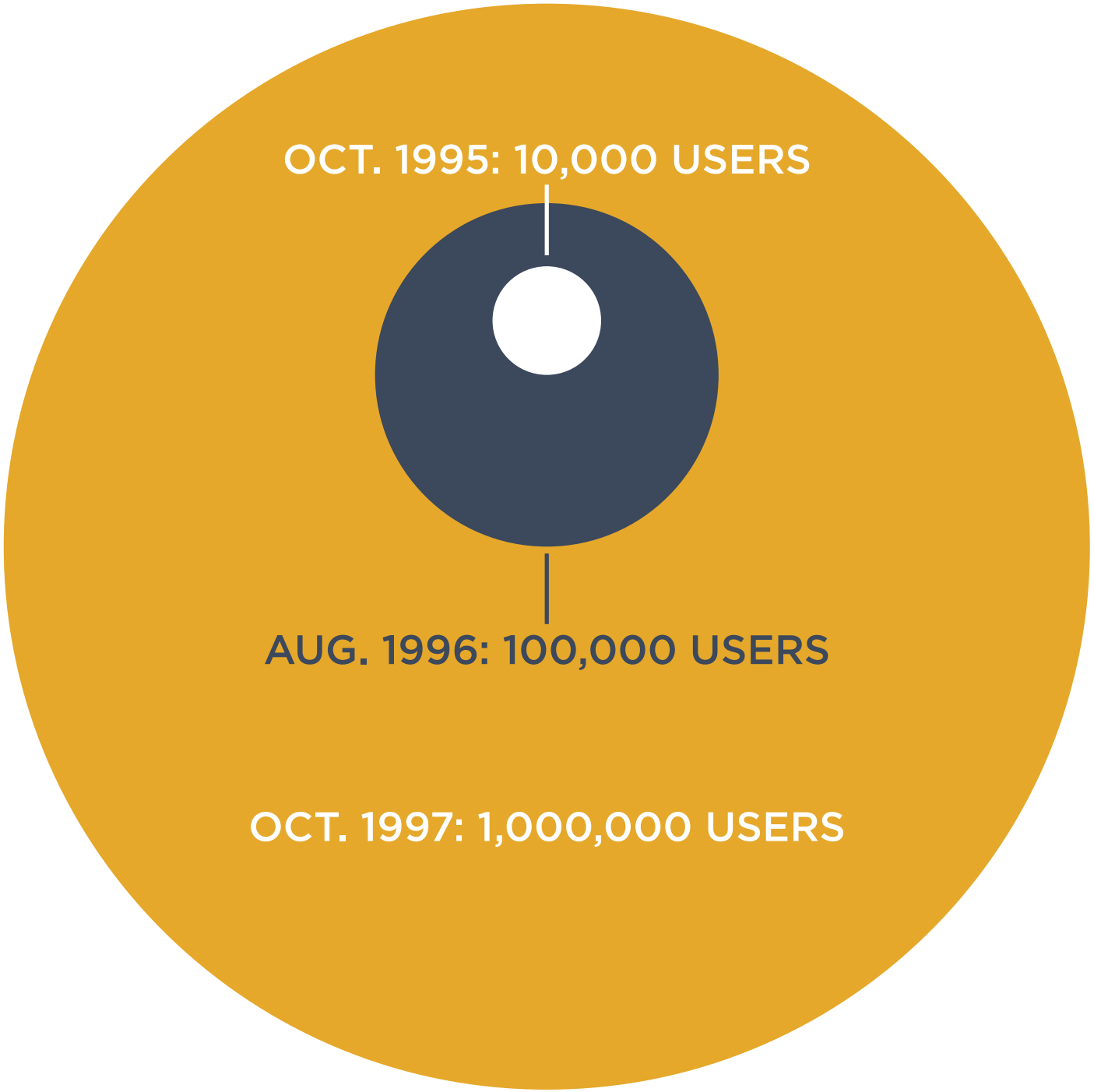
Why is a historian presenting a paper with such a title?

We have a problem facing our collective cultural heritage


Liberal.ca - Liberal Party
wayback.archive-it.org/227/20051004191404/http://liberal.ca/default_e.aspx
You are viewing an archived web page, collected at the request of University of Toronto using Archive-It. This page was captured on 19:14:04 Oct 04, 2005, and is part of the Canadian Political Parties and Political Interest Groups collection. The information on this web page may be out of date. See All versions of this archived page.
Liberal
HOME THE TEAM THE PARTY ISSUES MEDIA CENTRE YOUR RIDING DONATE
ADDRESS BY PRIME MINISTER PAUL MARTIN
TAKE ACTION TODAY !
Volunteer
Join Today!
Donate Now
Your Excellencies, Honourable Members, Ladies and Gentlemen:
Let me begin by expressing, on behalf of all Canadians, our appreciation to the Right Honourable Adrienne Clarkson and John Ralston Saul. With warmth, intelligence, and wit, they have honoured this high office and made an indelible contribution to our nation.
Over the course of six years, Madame Clarkson recognized achievement, decorated bravery, bore witness to tragedy and grief, and encouraged the disadvantaged. She welcomed foreign visitors and eloquently explained before audiences abroad what it is that makes Canada special. She took great interest in our cities and towns, and especially the north. She traveled to more than 200 communities across Canada; in some of them, it was the first-ever visit by a representative of the Crown.
Stay Informed
GO
Top Stories
September 29, 2005
Statement by the Prime Minister on the retirement of John Hamm, Premier of Nova Scotia
September 28, 2005
Charity Barbecue Raises $125,000 for Hurricane Katrina Victims
September 27, 2005
Address by Prime Minister Paul Martin at the installation of the new Governor General
Complete List of Stories
Commissions
Young Liberals of Canada
National Women's Liberal Commission
Aboriginal Peoples' Commission
Senior Liberals Commission


Cigardude's Smoking Room - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http://www.geocities.com/NapaValley/1070/
The Smoking Room
Your Choices
Cigars
Wine
Beer
Links
Home
Welcome to the Cigar Dude's Smoking Room
You are visitor number
since June 5, 1996
The main purpose of this page is to give me a forum to voice my views and opinions on cigars, good beer and fine wine. It's also a pretty good way for me to learn HTML. This page was first created on May 8, 1996 and will take some time to evolve, so if you are into cigars you might want to check back every once in a while to see what's up. It is always nice to know what other people
Welcome to my home page, devoted to some of the finer pleasures in life; good cigar
Start
Cigardude's Smoking ...
00:36

GEOCITIES USERS:




CONGRESSIONAL ARCHIVES
THOMAS P. O'NEILL PAPERS
PARTY LEADERSHIP / ADMINISTRATIVE FILES
Democratic Party
BOX 29
JOHN J. BURNS LIBRARY
BOSTON COLLEGE
374

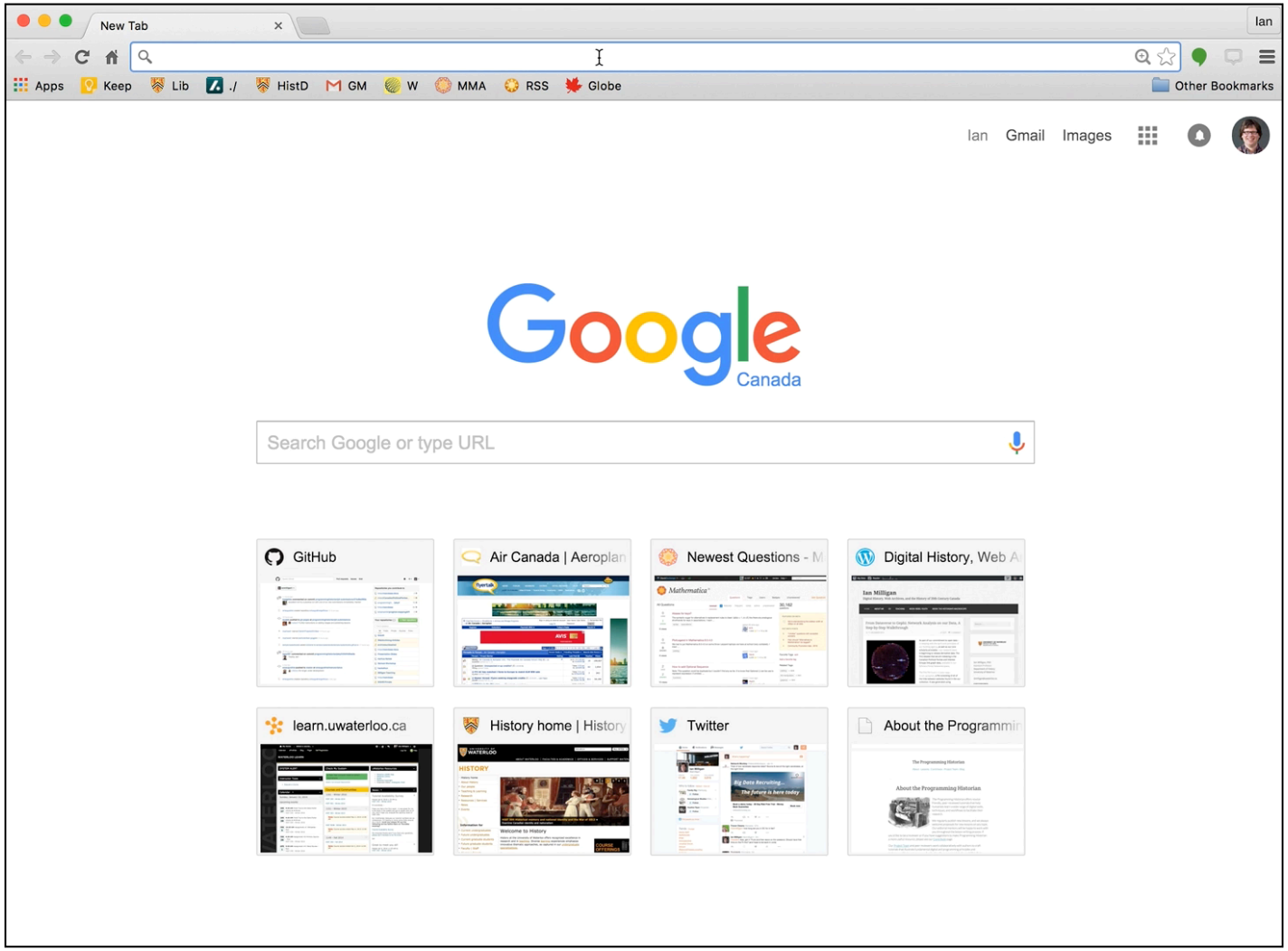
Scarcity

~~Scarcity~~

Abundance

Historians are largely unprepared to engage with the quantity of digital sources that will fundamentally transform their trade.

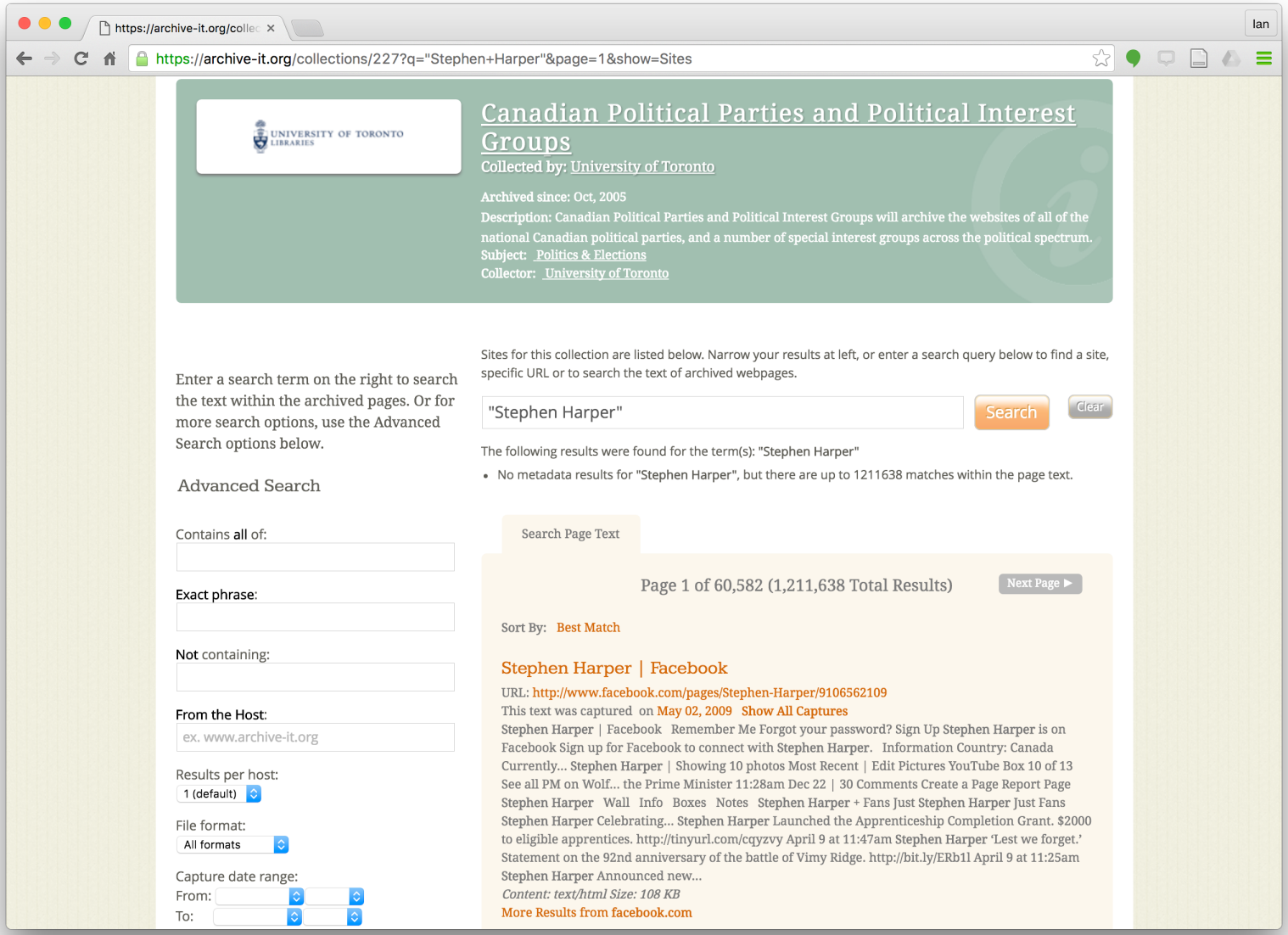
The Problem



This won't be enough!

… but what will our search engines look like?

Standard SERP Inadequate



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

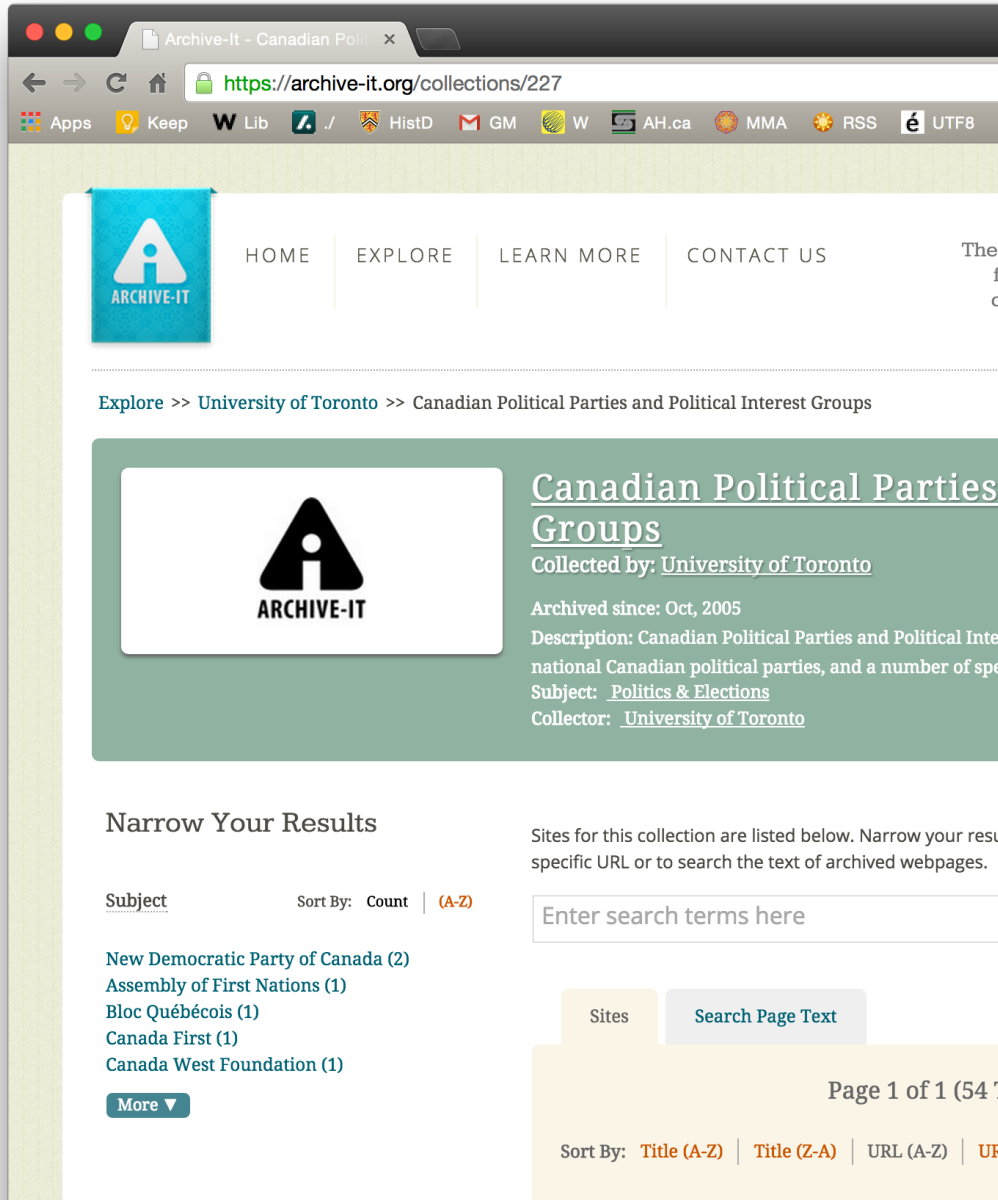
Schneiderman's mantra (1996)

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...

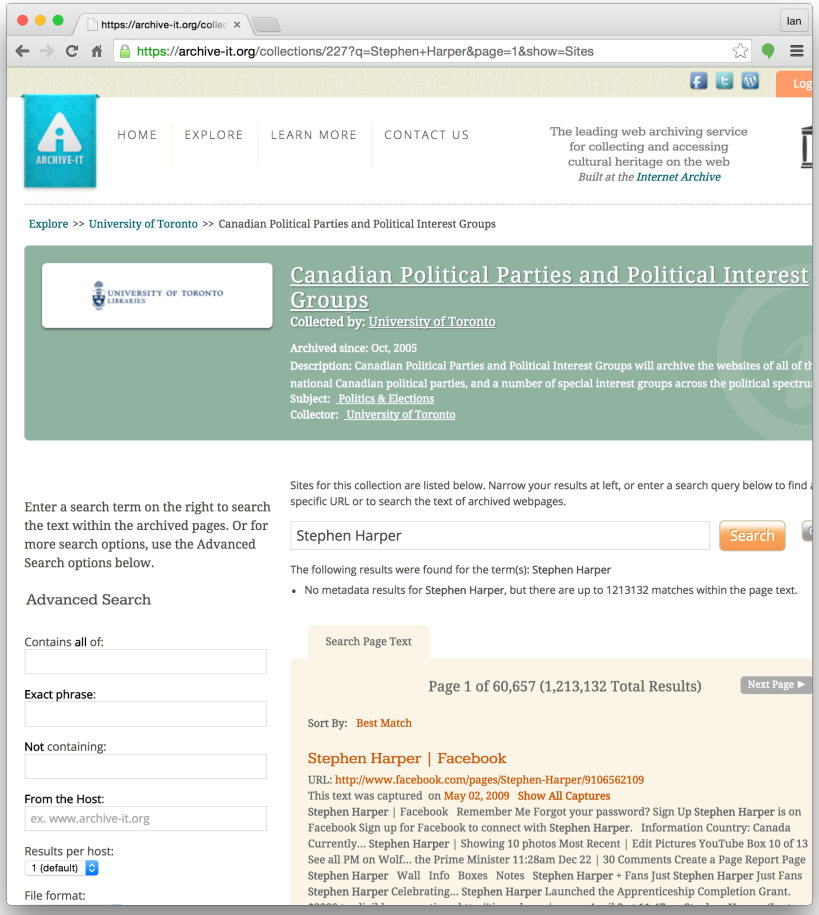


Canadian Political Parties & Political Interest Group Collection

- 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!



WebArchives.ca

Canadian Political Parties & Political Interest Group Collection

- **WebArchive-Discovery - https://github.com/ukwa/webarchive-discovery** (index to Solr, ingests ARCs/WARCs)
- **Shine - https://github.com/ukwa/shine** (frontend to Solr)
- **Warcbase - http://warcbase.org/** (contains Solr hadoop indexer)
- **Play Framework - https://www.playframework.com/**

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

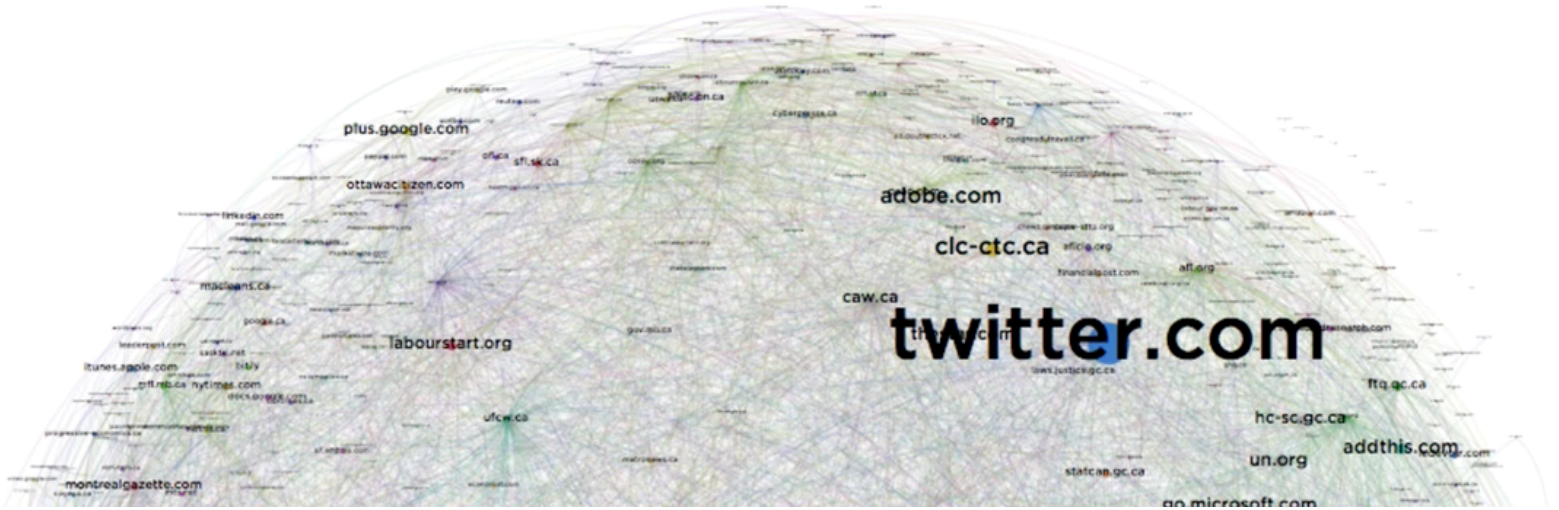
On this website, you can search web archived content from 50 political parties and political interest groups, from October 2005 to March 2015.

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.

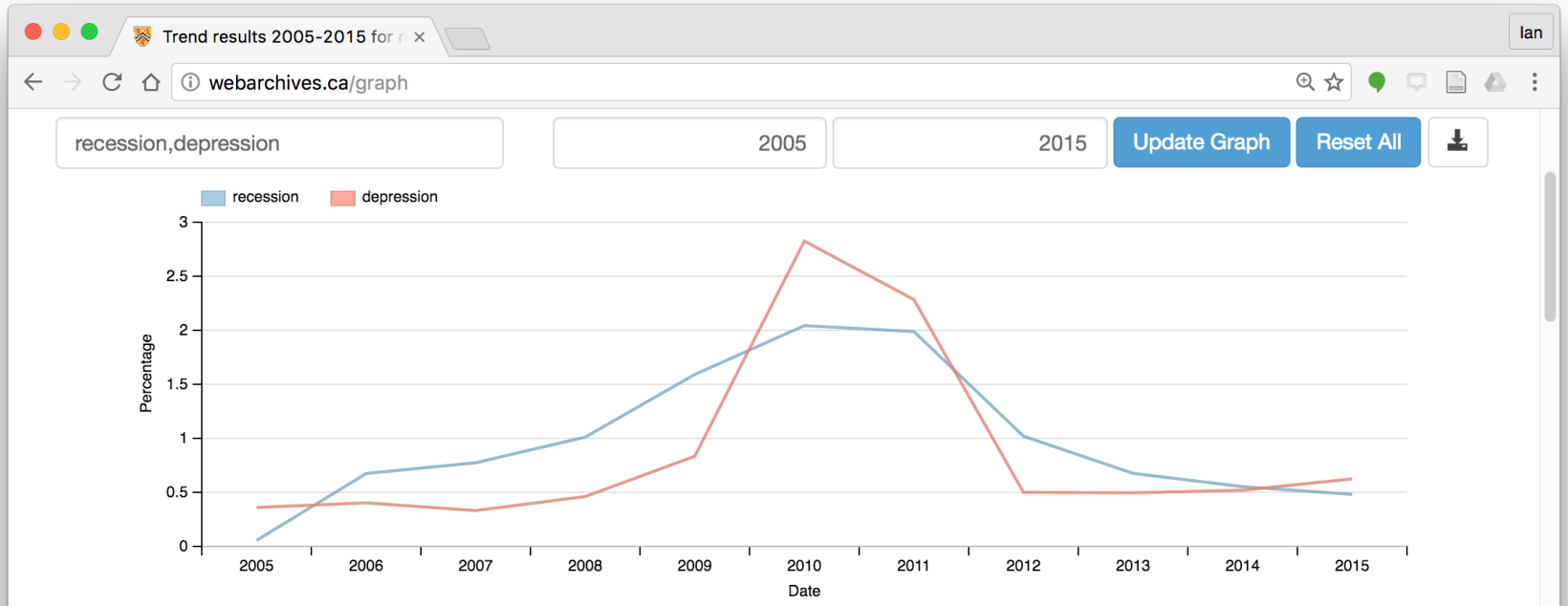
Options include:

- **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



Some Sample Queries to Get Started With:

- "Public Transit": Liberals/NDP/Conservatives. We use quotation marks to search on the exact phrase.
- "Tar Sands": Amongst the three major parties, who uses and subsequently stops using this phrase for Canada's Oil Sands?
- "Climate Change": Similar to the above two queries, we can see this issue rise and possibly wane in public interest.

Feel free to take these examples and adapt them to your own purposes. If you find something interesting, please let us know!

Found 100 samples matching 'recession' from 2010.

Matching Text	Link
recession , and if persisted in, a depression. This methodhits eve…	canadianactionparty.ca
. eeds forty percent of incomeand is rising. Today, with a recession starting	
prevent a recession at home rather than support the value of the dollar abr…	
…mblay=1064%5D" target="_blank">A Slowdown or a//: Recession in the U.S. in 2008?), we are	
…lemented," said James."For families struggling with the recession a	ndp.ca
..the financial burden we must live in today in this time of recession . I haven't	ndp.ca

Five Things We Learned

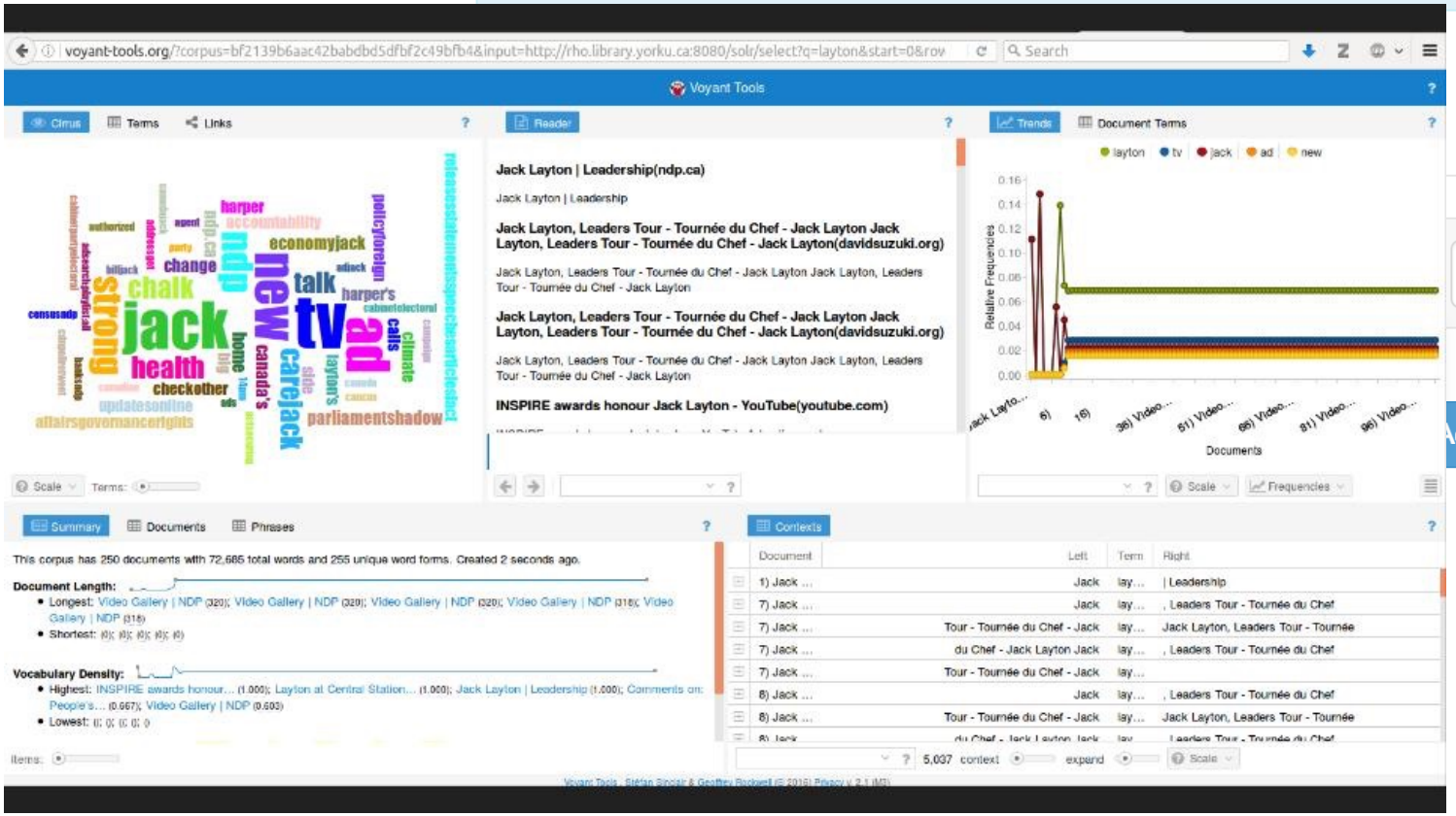
- Political parties delete content
- User-generated comments were more common in political parties
- Absences can be more informative than presences
- We can see the rise/fall of prominent people
- Enabling user access is truly transformative

Next Steps

- Moving to Blacklight;
- Integrating up to 25 different institutions; 150 collections into the index (currently signing MOUs);
- Developing better APIs;
- Exposing prioritization – transparency vs. quality, etc.

Next Steps

- Integration with Voyant-Tools, digital humanities suite

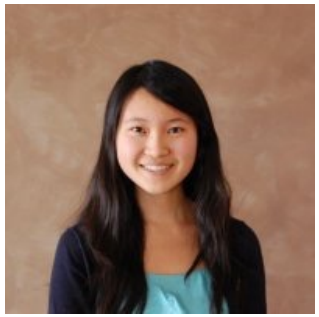


Our overriding mantra:
Make everything transparent,
work in interdisciplinary
teams

Interdisciplinary work at Waterloo/York

Historians

Computer Scientists



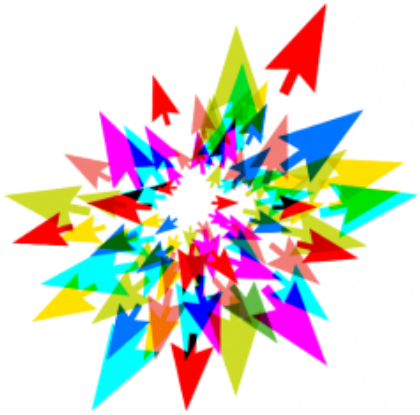
Librarians

Politics

Because, as I hope I have shown today..

it's worth it.

compute canada | calcul canada



Arts & Humanities Research Council

Thanks!

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Ontario

Andrew Jackson
Web Archiving Technical Lead
@anjacks0n

Jimmy Lin
Professor
@lintool

Ian Milligan
Assistant Professor
@ianmilligan1

Nick Ruest
Digital Assets Librarian
@ruebot

BRITISH LIBRARY

British Library

UNIVERSITY OF WATERLOO
FACULTY OF MATHEMATICS
David R. Cheriton School of Computer Science

UNIVERSITY OF WATERLOO
FACULTY OF ARTS
Department of History

YORK UNIVERSITÉ UNIVERSITY