See a little Warclight:

Building an open-source web archive portal with Project Blacklight

Nick Ruest (York University)
Ian Milligan (University of Waterloo)

IIPC WAC 2019 - Zagreb

Plan for The Talk

- Introduction
- Background
- Warclight
- Implementation at scale
- Next steps

Background

One perspective

Standing on Shoulders of Giants

- Analytical Access to the Domain Dark Archive
  - webarchive-discovery
- Big UK Domain Data for the Arts and Humanities
  - Shine
- Ruby on Rails
- Project Blacklight (UVa)
  - Library Catalogues
  - Panama Papers
  - GeoBlacklight
  - Arclight
  - Warclight

R+L = J

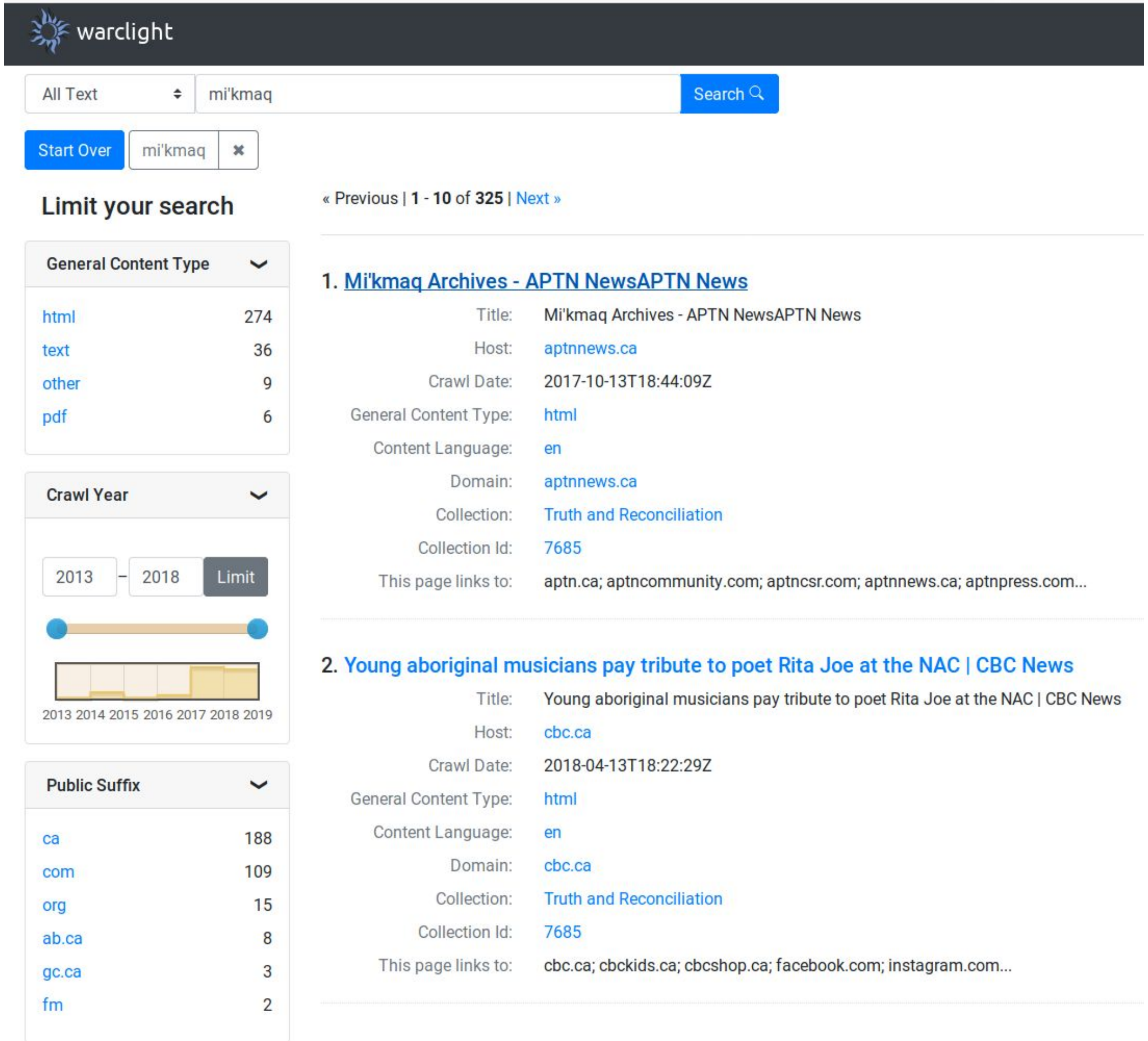
(Solr + webarchive-discovery) + Project Blacklight = Warclight

Warclight

What is it?

- Rails Engine
- Tight integration with webarchive-discovery
    - Collection, and Collection-ID
    - Institution
    - Resource name
- Features
    - Field search
    - Faceted search
    - Memento Replay URL check
    - Live web check
    - Thumbnail display
    - Plug-in architecture
        - Advanced Search
        - Range limit

😱 😱 Live Demo 😱 😱



All Text ⇕ | mi'kmaq | Search 🔍

Start Over | mi'kmaq ✖

« Previous | **1** - **10** of **325** | Next »

Limit your search

General Content Type

html	274
text	36
other	9
pdf	6

Crawl Year

2013 – 2018 Limit

Public Suffix

ca	188
com	109
org	15
ab.ca	8
gc.ca	3
fm	2

1. Mi'kmaq Archives - APTN NewsAPTN News

Title:	Mi'kmaq Archives - APTN NewsAPTN News
Host:	aptnnews.ca
Crawl Date:	2017-10-13T18:44:09Z
General Content Type:	html
Content Language:	en
Domain:	aptnnews.ca
Collection:	Truth and Reconciliation
Collection Id:	7685
This page links to:	aptn.ca; aptncommunity.com; aptncsr.com; aptnnews.ca; aptnpress.com...

2. Young aboriginal musicians pay tribute to poet Rita Joe at the NAC | CBC News

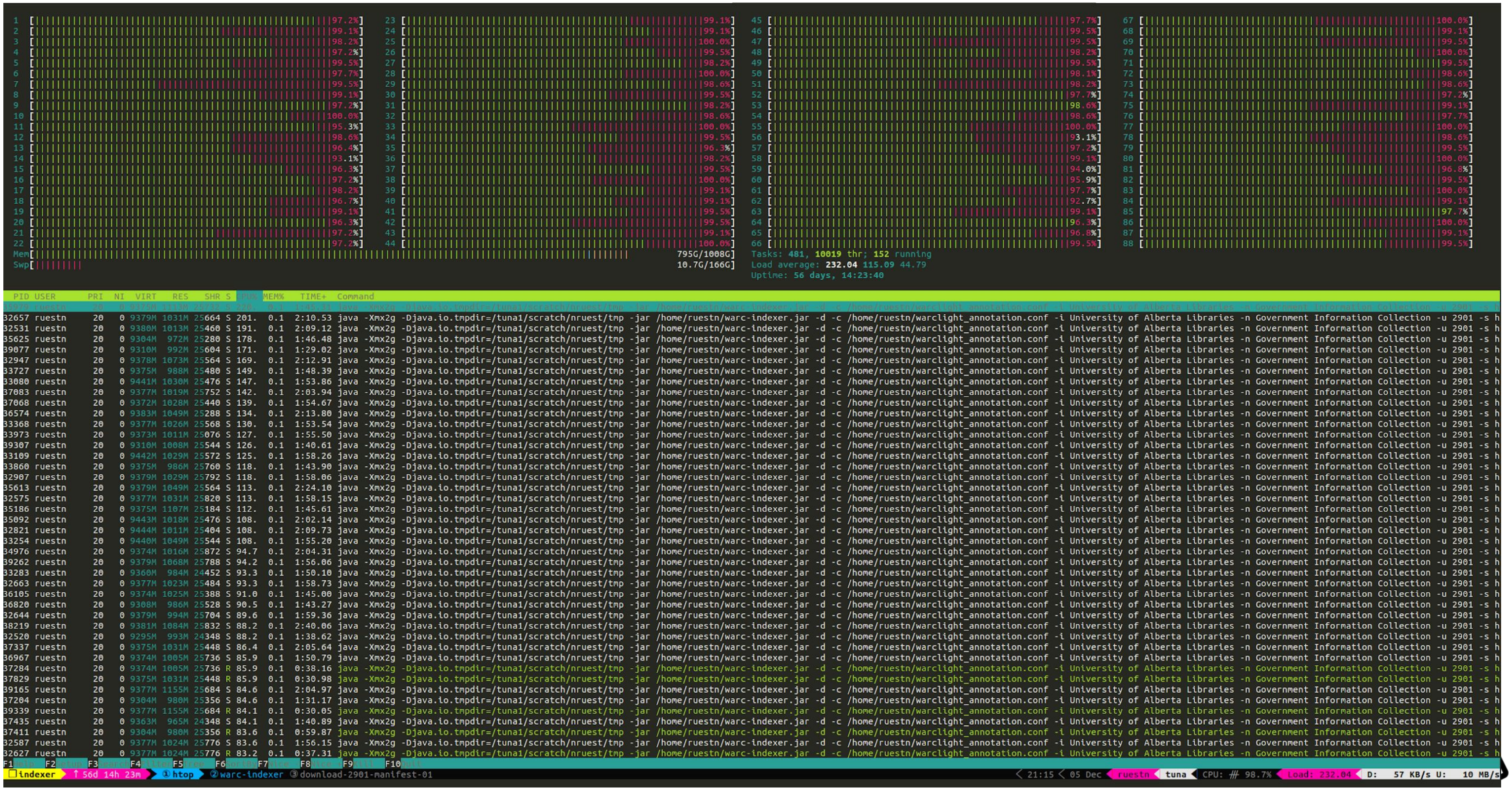
Title:	Young aboriginal musicians pay tribute to poet Rita Joe at the NAC | CBC News
Host:	cbc.ca
Crawl Date:	2018-04-13T18:22:29Z
General Content Type:	html
Content Language:	en
Domain:	cbc.ca
Collection:	Truth and Reconciliation
Collection Id:	7685
This page links to:	cbc.ca; cbckids.ca; cbcshop.ca; facebook.com; instagram.com...

Implementation at Scale

WALK

- 6 Canadian Universities
- Archive-It Partners
- 30T of WARCs
- 1+ Billion Solr docs
- ~1-2 months to download and index

Indexing



```
  1  [||||||||||||||||||||||||||||||||||||||||97.2%]   23 [||||||||||||||||||||||||||||||||||||||||99.1%]   45 [||||||||||||||||||||||||||||||||||||||||97.7%]   67 [|||||||||||||||||||||||||||||||||||||||100.0%]
  2  [||||||||||||||||||||||||||||||||||||||||99.1%]   24 [||||||||||||||||||||||||||||||||||||||||99.1%]   46 [||||||||||||||||||||||||||||||||||||||||99.5%]   68 [||||||||||||||||||||||||||||||||||||||||99.1%]
  3  [||||||||||||||||||||||||||||||||||||||||98.2%]   25 [|||||||||||||||||||||||||||||||||||||||100.0%]   47 [||||||||||||||||||||||||||||||||||||||||99.5%]   69 [||||||||||||||||||||||||||||||||||||||||99.5%]
  4  [||||||||||||||||||||||||||||||||||||||||97.2%]   26 [||||||||||||||||||||||||||||||||||||||||99.5%]   48 [||||||||||||||||||||||||||||||||||||||||98.2%]   70 [|||||||||||||||||||||||||||||||||||||||100.0%]
  5  [||||||||||||||||||||||||||||||||||||||||99.5%]   27 [||||||||||||||||||||||||||||||||||||||||98.2%]   49 [||||||||||||||||||||||||||||||||||||||||99.5%]   71 [||||||||||||||||||||||||||||||||||||||||99.5%]
  6  [||||||||||||||||||||||||||||||||||||||||97.7%]   28 [|||||||||||||||||||||||||||||||||||||||100.0%]   50 [||||||||||||||||||||||||||||||||||||||||98.1%]   72 [||||||||||||||||||||||||||||||||||||||||98.6%]
  7  [||||||||||||||||||||||||||||||||||||||||99.5%]   29 [||||||||||||||||||||||||||||||||||||||||98.6%]   51 [||||||||||||||||||||||||||||||||||||||||98.2%]   73 [||||||||||||||||||||||||||||||||||||||||98.6%]
  8  [||||||||||||||||||||||||||||||||||||||||99.1%]   30 [||||||||||||||||||||||||||||||||||||||||99.5%]   52 [||||||||||||||||||||||||||||||||||||||||97.7%]   74 [||||||||||||||||||||||||||||||||||||||||97.2%]
  9  [||||||||||||||||||||||||||||||||||||||||97.2%]   31 [||||||||||||||||||||||||||||||||||||||||98.2%]   53 [||||||||||||||||||||||||||||||||||||||||98.6%]   75 [||||||||||||||||||||||||||||||||||||||||99.1%]
  10 [|||||||||||||||||||||||||||||||||||||||100.0%]   32 [||||||||||||||||||||||||||||||||||||||||98.6%]   54 [||||||||||||||||||||||||||||||||||||||||98.6%]   76 [||||||||||||||||||||||||||||||||||||||||97.7%]
  11 [||||||||||||||||||||||||||||||||||||||||95.3%]   33 [|||||||||||||||||||||||||||||||||||||||100.0%]   55 [|||||||||||||||||||||||||||||||||||||||100.0%]   77 [|||||||||||||||||||||||||||||||||||||||100.0%]
  12 [||||||||||||||||||||||||||||||||||||||||98.6%]   34 [||||||||||||||||||||||||||||||||||||||||99.5%]   56 [||||||||||||||||||||||||||||||||||||||||93.1%]   78 [||||||||||||||||||||||||||||||||||||||||98.6%]
  13 [||||||||||||||||||||||||||||||||||||||||96.4%]   35 [||||||||||||||||||||||||||||||||||||||||96.3%]   57 [||||||||||||||||||||||||||||||||||||||||97.2%]   79 [||||||||||||||||||||||||||||||||||||||||99.5%]
  14 [||||||||||||||||||||||||||||||||||||||||93.1%]   36 [||||||||||||||||||||||||||||||||||||||||98.2%]   58 [||||||||||||||||||||||||||||||||||||||||99.1%]   80 [|||||||||||||||||||||||||||||||||||||||100.0%]
  15 [||||||||||||||||||||||||||||||||||||||||96.3%]   37 [||||||||||||||||||||||||||||||||||||||||99.5%]   59 [||||||||||||||||||||||||||||||||||||||||94.0%]   81 [||||||||||||||||||||||||||||||||||||||||96.8%]
  16 [||||||||||||||||||||||||||||||||||||||||97.2%]   38 [|||||||||||||||||||||||||||||||||||||||100.0%]   60 [||||||||||||||||||||||||||||||||||||||||95.9%]   82 [||||||||||||||||||||||||||||||||||||||||99.5%]
  17 [||||||||||||||||||||||||||||||||||||||||98.2%]   39 [||||||||||||||||||||||||||||||||||||||||99.1%]   61 [||||||||||||||||||||||||||||||||||||||||97.7%]   83 [|||||||||||||||||||||||||||||||||||||||100.0%]
  18 [||||||||||||||||||||||||||||||||||||||||96.7%]   40 [||||||||||||||||||||||||||||||||||||||||99.1%]   62 [||||||||||||||||||||||||||||||||||||||||92.7%]   84 [||||||||||||||||||||||||||||||||||||||||99.1%]
  19 [||||||||||||||||||||||||||||||||||||||||99.1%]   41 [||||||||||||||||||||||||||||||||||||||||99.5%]   63 [||||||||||||||||||||||||||||||||||||||||99.1%]   85 [||||||||||||||||||||||||||||||||||||||||97.7%]
  20 [||||||||||||||||||||||||||||||||||||||||96.3%]   42 [||||||||||||||||||||||||||||||||||||||||99.5%]   64 [||||||||||||||||||||||||||||||||||||||||96.3%]   86 [|||||||||||||||||||||||||||||||||||||||100.0%]
  21 [||||||||||||||||||||||||||||||||||||||||97.2%]   43 [||||||||||||||||||||||||||||||||||||||||99.1%]   65 [||||||||||||||||||||||||||||||||||||||||96.8%]   87 [||||||||||||||||||||||||||||||||||||||||99.1%]
  22 [||||||||||||||||||||||||||||||||||||||||97.2%]   44 [|||||||||||||||||||||||||||||||||||||||100.0%]   66 [||||||||||||||||||||||||||||||||||||||||99.5%]   88 [||||||||||||||||||||||||||||||||||||||||99.5%]
  Mem[||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||          795G/1008G]   Tasks: 481, 10019 thr; 152 running
  Swp[||||||||                                                                                  10.7G/166G]   Load average: 232.04 115.09 44.79
                                                                                                              Uptime: 56 days, 14:23:40

  PID USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
35979 ruestn     20   0 9375M 1113M 25732 S 220.  0.1  1:45.31 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32657 ruestn     20   0 9379M 1031M 25664 S 201.  0.1  2:10.53 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32531 ruestn     20   0 9380M 1013M 25460 S 191.  0.1  2:09.12 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
35625 ruestn     20   0 9304M  972M 25280 S 178.  0.1  1:46.48 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
39077 ruestn     20   0 9310M  992M 25604 S 171.  0.1  1:29.02 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32947 ruestn     20   0 9378M 1073M 25564 S 169.  0.1  2:12.91 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33727 ruestn     20   0 9375M  988M 25480 S 149.  0.1  1:48.39 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33080 ruestn     20   0 9441M 1030M 25476 S 147.  0.1  1:53.86 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37083 ruestn     20   0 9377M 1019M 25752 S 142.  0.1  2:03.94 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37068 ruestn     20   0 9372M 1028M 25440 S 139.  0.1  1:54.67 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
36574 ruestn     20   0 9383M 1049M 25288 S 134.  0.1  2:13.80 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33368 ruestn     20   0 9377M 1026M 25568 S 130.  0.1  1:53.54 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33973 ruestn     20   0 9373M 1011M 25076 S 127.  0.1  1:55.50 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
39307 ruestn     20   0 9310M 1008M 25544 S 126.  0.1  1:40.61 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33109 ruestn     20   0 9442M 1029M 25572 S 125.  0.1  1:58.26 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33860 ruestn     20   0 9375M  986M 25760 S 118.  0.1  1:43.90 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32907 ruestn     20   0 9379M 1029M 25792 S 118.  0.1  1:58.06 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
35613 ruestn     20   0 9379M 1049M 25564 S 113.  0.1  2:24.10 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32575 ruestn     20   0 9377M 1031M 25820 S 113.  0.1  1:58.15 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
35186 ruestn     20   0 9375M 1107M 25184 S 112.  0.1  1:45.61 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
35092 ruestn     20   0 9443M 1018M 25476 S 108.  0.1  2:02.14 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32821 ruestn     20   0 9444M 1011M 25404 S 108.  0.1  2:09.73 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33254 ruestn     20   0 9440M 1049M 25544 S 108.  0.1  1:55.20 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
34976 ruestn     20   0 9374M 1016M 25872 S 94.7  0.1  2:04.31 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
39262 ruestn     20   0 9379M 1068M 25788 S 94.2  0.1  1:56.06 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
33283 ruestn     20   0 9360M  984M 24452 S 93.3  0.1  1:50.10 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32663 ruestn     20   0 9377M 1023M 25484 S 93.3  0.1  1:58.73 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
36105 ruestn     20   0 9374M 1025M 25388 S 91.0  0.1  1:45.00 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
36820 ruestn     20   0 9308M  986M 25528 S 90.5  0.1  1:43.27 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32644 ruestn     20   0 9379M  994M 25704 S 89.6  0.1  1:59.36 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
38219 ruestn     20   0 9381M 1084M 25832 S 88.2  0.1  2:40.06 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32520 ruestn     20   0 9295M  993M 24348 S 88.2  0.1  1:38.62 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37337 ruestn     20   0 9375M 1031M 25448 S 86.4  0.1  2:05.64 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
36967 ruestn     20   0 9374M 1005M 25736 S 85.9  0.1  1:50.79 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37284 ruestn     20   0 9374M 1005M 25736 R 85.9  0.1  0:38.16 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37829 ruestn     20   0 9375M 1031M 25448 R 85.9  0.1  0:30.98 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
39165 ruestn     20   0 9377M 1155M 25684 S 84.6  0.1  2:04.97 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37204 ruestn     20   0 9304M  980M 25356 S 84.6  0.1  1:31.17 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
39339 ruestn     20   0 9377M 1155M 25684 R 84.1  0.1  0:30.05 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37435 ruestn     20   0 9363M  965M 24348 S 84.1  0.1  1:40.89 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
37411 ruestn     20   0 9304M  980M 25356 R 83.6  0.1  0:59.87 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32587 ruestn     20   0 9377M 1024M 25776 S 83.6  0.1  1:56.15 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
32627 ruestn     20   0 9377M 1024M 25776 R 83.2  0.1  0:37.31 java -Xmx2g -Djava.io.tmpdir=/tuna1/scratch/nruest/tmp -jar /home/ruestn/warc-indexer.jar -d -c /home/ruestn/warclight_annotation.conf -i University of Alberta Libraries -n Government Information Collection -u 2901 -s h
F1Help  F2Setup F3SearchF4FilterF5Tree  F6SortByF7Nice -F8Nice +F9Kill  F10Quit
□indexer  ↑ 56d 14h 23m  ①htop  ②warc-indexer ③download-2901-manifest-01        < 21:15 < 05 Dec  ruestn  tuna  CPU: ## 98.7%  Load: 232.04  D:   57 KB/s U:   10 MB/s
```

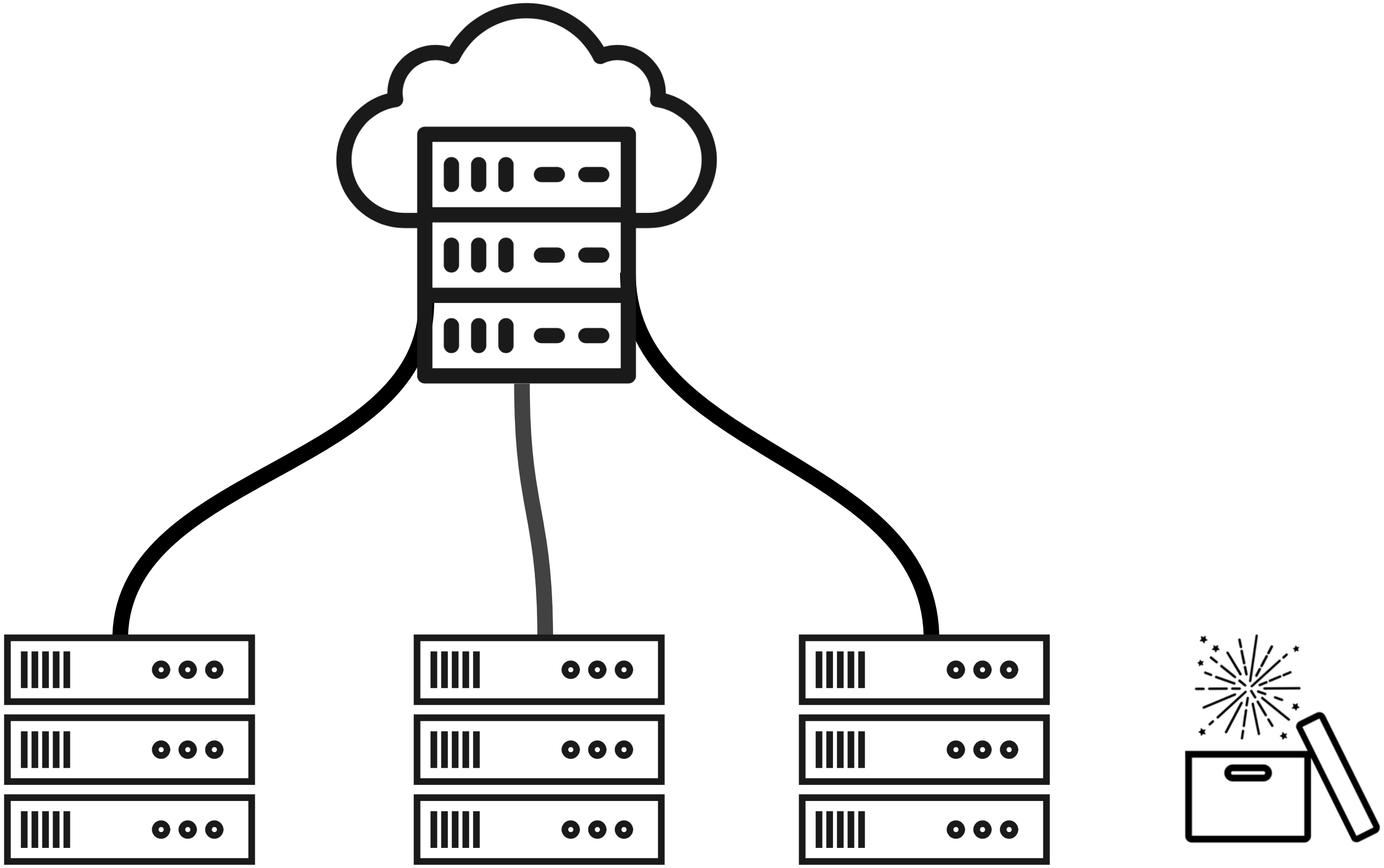
Warclight apps

- warclight demo - warclight.archivesunleashed.org (2,775,723)
- University of Toronto - utoronto.archivesunleashed.org (691,803,697)
- University of Alberta - ualberta.archivesunleashed.org (331,924,814)
- University of Victoria - uvic.archivesunleashed.org (18,873,434)
- University of Winnipeg - uwinnipeg.archivesunleashed.org (30,793,376)
- Dalhousie University - dalhousie.archivesunleashed.org (736,023)
- Simon Fraser University - sfu.archivesunleashed.org (787,991)
- GeoCities - geocities.archivesunleashed.org (316,694,661)
- WALK (federated) - walk.archivesunleashed.org (1,077,695,058)

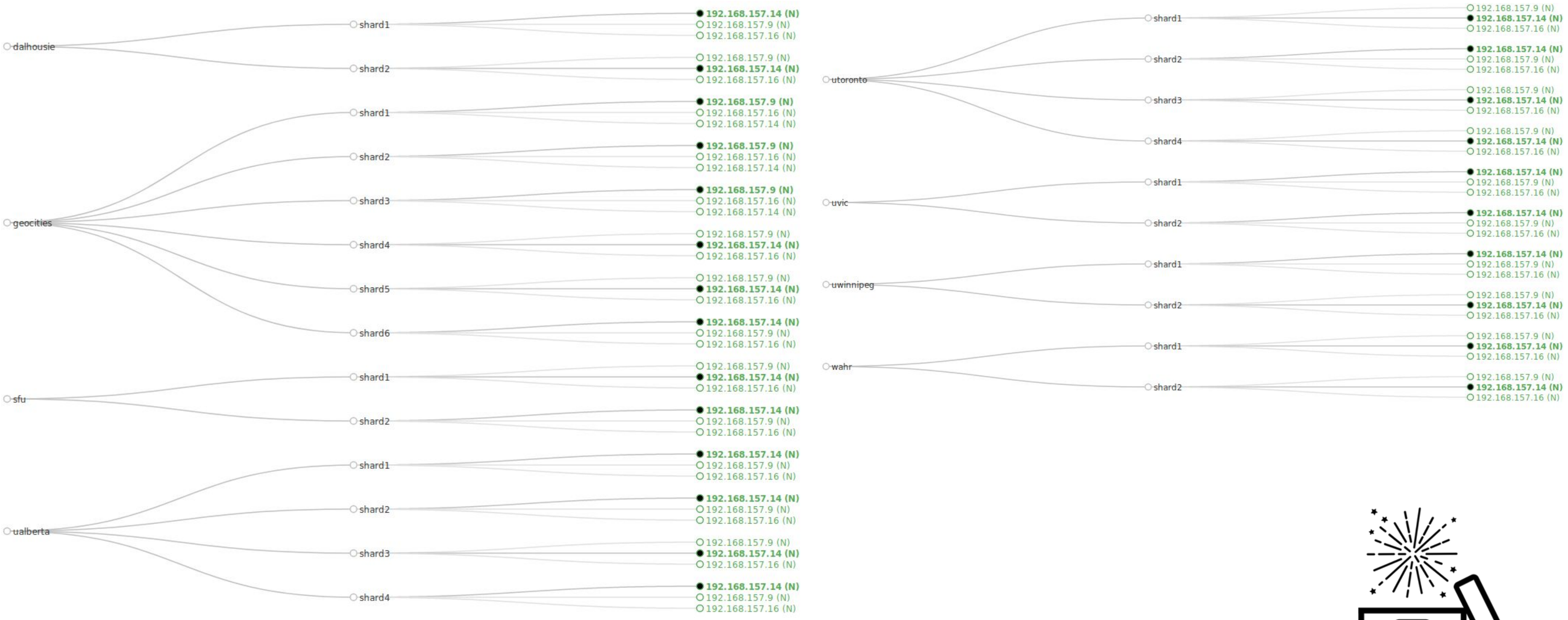
Components

APACHE
HTTP SERVER PROJECT
HAPROXY
PHUSION
ZOO
Solr
RAILS

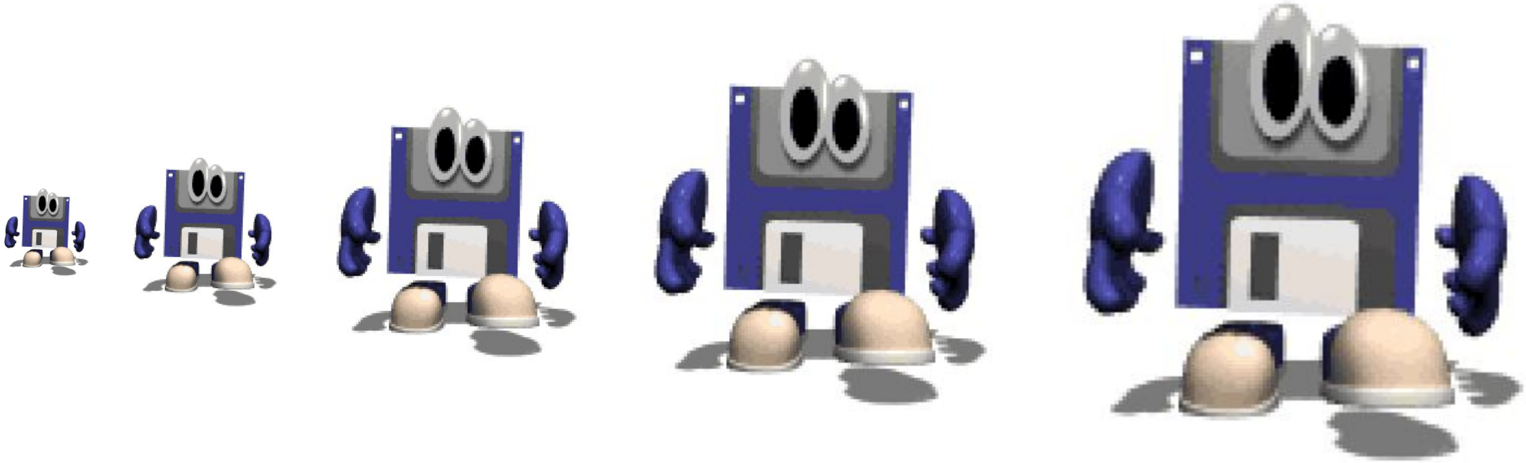
System diagram



SolrCloud Collections



Implementation



Facet caching:

- https://issues.apache.org/jira/browse/SOLR-13132
- https://github.com/magibney/solr-facet-cache

Solr index:

- 😱
- 💸

Next Steps

Discussion

- Another framework for webarchive-discovery
- Ruby on Rails vs Play Framework
- Parity features with Shine?
  - Concordance
  - Trend search
  - Blacklight plugins
- Growing the community - is there one?
- Reality of time and Archives Unleashed priorities
- Is this even sustainable?? (Solr index size)

We look forward to your questions and thoughts.

Thank you!

- Toke Eskildsen, Thomas Egense
- Andy Jackson, Gil Hoggarth
- Chris Beer, Jesse Keck, Justin Coyne (Blacklight Community)

Thanks to our supporters!

THE ANDREW W. MELLON FOUNDATION

compute canada | calcul canada

START SMART LABS

YORK UNIVERSITÉ UNIVERSITY

UNIVERSITY OF WATERLOO

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Links

- archivesunleashed.org/warclight
- archivesunleashed.org
- cloud.archivesunleashed.org
- github.com/archivesunleashed
- slack.archivesunleashed.org
- news.archivesunleashed.org
- twitter.com/unleasharchives