See a little Warclight: Building an open-source web archive portal with Project Blacklight

Nick Ruest (York University)  
Ian Milligan (University of Waterloo)
Plan for The Talk

- Introduction
- Background
- Warclight
- Implementation at scale
- Next steps
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 Reply URL check
  - Live web check
  - Thumbnail display
  - Plug-in architecture
    - Advanced Search
    - Range limit
Implementation at Scale
- 6 Canadian Universities
- Archive-It Partners
- 30T of WARCs
- 1+ Billion Solr docs
- ~1-2 months to download and index
Indexing
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 - uwinipeg.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

HAProxy

Solr

Rails
System diagram
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
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!
Links

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