Walk Before You Run

Prerequisites to Linked Data

Kenning Arlitsch
Dean of the Library
@kenning_msu
Linked Data applications will not matter if search engines can’t first find library websites and repositories, crawl them, and understand the metadata provided.
Agenda

Traditional SEO (Search Engine Optimization)
  – Optimize hardware, software, websites, metadata

Semantic Web Optimization
  – Semantic Identity
  – Schema.org Projects at MSU
    • Using a vocabulary understood by search engines
    • Improve machine comprehension
Funded Research

• 2011-2014
  – “Getting Found: Search Engine Optimization for Digital Repositories”

• 2014-2017
  – Partners
    • OCLC Research
    • Association of Research Libraries
    • University of New Mexico
SEARCH ENGINE OPTIMIZATION

Part 1 of 3
SEO Building Blocks

• Priority 1 – Increase Reach
  – Get objects indexed by search engines

• Priority 2 – Increase Visibility in SERP
  – Provide robust descriptive content

• Priority 3 – Get Relevant
  – Increase click-through rates (CTR)
Why it Matters
Where College Students Begin Research

Americans submit 18 billion search queries to search engines each month*

- 12 billion to Google sites (67%)
- 3.5 billion to Microsoft sites (19%)
- 1.8 billion to Yahoo! Sites (10%)
- 355 million to Ask Networks (.02%)
- 224 million to AOL, Inc. (.01%)

How much of that traffic is directed to our libraries?

Our Research Inspiration

• Ten years building digital libraries
  – Mountain West Digital Library
  – Utah Digital Newspapers
  – Western Waters Digital Library
  – Western Soundscape Archive

• Were they being used?
Basic SEO improved Utah’s collection accessibility in Google...

Google Index Ratio - All Collections*

- Google Index Ratio = URLs submitted / URLs Indexed by Google
- ~150 collections containing ~170,000 URLs (07/2010) and ~170 collections containing ~282,000 URLs (12/2013)
...Producing significant increases in the average number of page views per day...
...resulting in more referrals and visitors

12 week comparison 2010 vs. 2012

Referring Domain: Search Results For google

<table>
<thead>
<tr>
<th>Domain</th>
<th>Custom View: 2/6/12 - 4/29/12 Visits</th>
<th>Custom View: 2/1/10 - 4/25/10 Visits</th>
<th>% Change Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>google.com</td>
<td>51,694</td>
<td>8,959</td>
<td>477.01%</td>
</tr>
<tr>
<td>google.co.uk</td>
<td>1,284</td>
<td>182</td>
<td>605.49%</td>
</tr>
<tr>
<td>google.ca</td>
<td>1,203</td>
<td>415</td>
<td>189.88%</td>
</tr>
<tr>
<td>google.it</td>
<td>670</td>
<td>38</td>
<td>1,663.16%</td>
</tr>
<tr>
<td>google.co.in</td>
<td>602</td>
<td>68</td>
<td>785.29%</td>
</tr>
<tr>
<td>google.fr</td>
<td>475</td>
<td>35</td>
<td>1,257.14%</td>
</tr>
<tr>
<td>google.es</td>
<td>466</td>
<td>26</td>
<td>1,692.31%</td>
</tr>
<tr>
<td>google.com.au</td>
<td>463</td>
<td>95</td>
<td>387.37%</td>
</tr>
<tr>
<td>google.de</td>
<td>441</td>
<td>88</td>
<td>401.14%</td>
</tr>
<tr>
<td>google.com.br</td>
<td>408</td>
<td>29</td>
<td>1,660.96%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>63,637</td>
<td>10,559</td>
<td><strong>502.68%</strong></td>
</tr>
</tbody>
</table>
Technical Barriers to SE Crawlers

• Website Design
  – Graphics
  – Confusing site hierarchies and paths

• Slow servers

• CMS often lack canonical links

• Metadata
  – Schema not understood by SE
  – Not unique
  – Inconsistent/inaccurate
Challenge is presenting structured data SE’s can identify, parse and digest.

*Human Readable*

Thanks for nothing: changes in income and labor force participation for never-married mothers since 1982
NH Wolfinger, M McKeever - ... Annual Meeting; 2006 Aug 11-14 ...; 2006 - cdm6gs.lib.utah.edu
Abstract This study examines whether the changing social and economic characteristics of women who give birth out of wedlock have led to higher family incomes. Using Current Population Survey data collected between 1982 and 2002, we find that never-married mothers...
Related articles All 2 versions Cite More▼

Thanks for Nothing: Changes in Income and Labor Force Participation for Never-Married Mothers since 1982
M McKeever, NH Wolfinger - 2007 - archiv.soc.cas.cz
ABSTRACT This study examines whether the changing social and economic characteristics of women who give birth out of wedlock have led to higher family incomes. Using Current Population Survey data collected between 1982 and 2002, we find that never-married mothers...
Related articles All 2 versions Cite More▼

Google Scholar can read and understand!
Organizational/Cultural Themes

• Traditional SEO is an afterthought
• Librarians think too small re potential traffic
• Organizational communication is poor
• Analytics are usually poorly implemented
• Vendors are slow to catch on to SEO problems
  – Because we don’t demand it
How does a library justify SEO?

- Promote use of collections
- Accountability to funding organizations
- Improve institutional reputation
  - Increase citation rates
  - Attract research funding
  - Attract students and faculty
Recommended SEO Process

• Institutionalize SEO
  – Strategic Plan/driven by administration
  – Accurate Measurement Tools

• Traditional SEO
  – Get Indexed = Index Ratio
  – Get Visible = Search Engine Results Page (SERP)

• Semantic Web Optimization
  – Get Relevant = Click Through Rates (CTR)
Part 2 of 3

SEMANTIC WEB OPTIMIZATION
Semantic Identity

Library organizations are poorly defined and represented on the Semantic Web...
Google’s Knowledge Graph

“A knowledge base ... to enhance search results with semantic-search information gathered from a wide variety of sources”

President Olafur Ragnar Grimsson visited Cornell in late November. Among other stops on campus, he came to the library to tour Cornell’s renowned Fiske...

Digital Library
The Alfredo Montalvo Bolivian Digital Pamphlets Collection is ...

My Account
... Services • Ask a Librarian • Help • My Account • Home; My ...

About Us
About Us. Welcome! In a time of rapid change, Cornell University ...

More results from cornell.edu »

Rare and Manuscript Collect...
Collections • Digital Collections • Carl A. Kroch Library • ...

Libraries and Hours
Libraries and Hours. View map of Cornell libraries. Show only ...

University Librarian
In a time of rapid change, Cornell University Library continues to ...

Map for Cornell University Library

Cornell University Library
www.library.cornell.edu
4 Google reviews

Cornell University, 201 Cornell Library
Ithaca, NY
(607) 255-4144
Google’s Perception of MSU Lib - 2012

Montana State University Library
www.lib.montana.edu/
Montana State University in Bozeman. Ask A Librarian ask a librarian. Ronnie Library, 1st floor. Friday, May 17: 10:00-5:00. Saturday, May 18: 10:00-1:00 ...

Articles & Research Databases :: Montana State University Library
www.lib.montana.edu/resources/
Articles & Research Databases. Home » Resources. By title. A B C D E F ...

Digital Collections - Montana State University Library
www.lib.montana.edu/digital/
MSU Library Digital Initiative collections can be full digital object retrieval or ...

Special Collections & Archives :: Montana State University Library
www.lib.montana.edu/archives/
Special Collections & Archives. The Merrill G. Buringame Special Collections ...

Contact Us :: Montana State University Library
www.lib.montana.edu/contact.php
Contact Us. Comments and Feedback. General Questions and Comments for ...

A-Z Site Index :: Montana State University Library
www.lib.montana.edu/sitemap.php
A-Z Site Index. Browse by page title: A B C D E F G H I J K L M N O P Q R ...

Montana State University (MSU) Library, Mobile
www.lib.montana.edu/m/
MSU Library (Mobile). Search; Databases. Hours; Ask a Librarian. About ...

Montana State University
Address: 1500 University Dr, Billings, Mt 59101
Phone: (406) 657-1662
Hours: Wednesday hours 7:30 am-10:00 pm - See all
Where does Google get its information?
Trusted Sources for Search Engines

• No Wikipedia presence?
  – You don’t exist as an important “entity” or “thing”
  – You exist as a string of (confusing) text

• Other influences on the Knowledge Graph
  – FreeBase (phasing out)
  – Wikidata
  – Google Places/Google My Business
  – Google+
The page "Montana state university library" does not exist. You can ask for it to be created, but consider checking the search results below to see whether the topic is already covered.

For search help, please visit Help:Searching.

Bozeman, Montana
County. Montana, United States. 61 dateform mdy in the southwestern part of the state. ... College, and in 1965 it became Montana State University ... 61 KB (8,314 words) - 03:40, 19 August 2013

Boise, Idaho (section Boise State University)
But following the creation of Montana Territory, Boise was made the ... Boise State University: Boise State University and its surrounding environs. ... 61 KB (8,349 words) - 21:43, 3 September 2013

Montana State University Billings
Montana State University Billings (or MSU Billings) is a state university ... Major structures of the main campus: Library, Liberal Arts Building ... 15 KB (1,985 words) - 20:21, 26 April 2013

Montana State University
Montana State University (MSU) is a public university located in Bozeman, Montana ... These included the 1946 Library Building (now Renne ... 65 KB (9,267 words) - 14:59, 27 August 2013

Carl W. McIntosh (section Montana State University)
He was named president of Montana State University in 1970. ... Wheels, and operated regular delivery of books from the Bozeman Public Library.
DBPedia entry - 2012

About: http://dbpedia.org/resource/Montana_State_University_Library
An Entity of Type: Thing, from Named Graph: http://dbpedia.org, within Data Space: dbpedia.org

No further information is available. (The requested entity is unknown)
Montana State University Library (MSU Library) is the academic library of Montana State University, Montana's land-grant university, in Bozeman, Montana, United States. It is the flagship library for all of Montana State University System's campuses. In 1978, the library was named the Roland R. Renne Library to honor the sixth president of the university. The library supports the research and information needs of Montana's students, faculty, and the Montana Extension Service.

Contents

1 History
2 Collections
   2.1 Digital collections
   2.2 ScholarWorks
   2.3 Special Collections and University Archives
3 Budget
4 Special events
5 Awards
6 Memberships
7 See also
8 References

History

In January 1894, about seven months after Montana College of Agriculture and Mechanic Arts was founded, the college began acquiring and housing a formal library collection for its students and faculty research use. For the first two years of the library's existence, students or instructors served part-time to provide library services. In 1896, Mabel Ruth Owens became the first full-time professional librarian to oversee the library's operations and collection development. In 1927, the library was moved to the second floor of Montana Hall.[2]

In 1949, the library, its collection and research services were moved to a newly constructed 6,894 square feet (638.3 m²) library facility.[3] After the transfer of nearly 100,000 volumes from various locations around campus the library officially opened its doors on January 9, 1950.[4]
Montana State University Library (MSU Library) is the academic library of Montana State University, Montana's land-grant university, in Bozeman, Montana, United States. It is the flagship library for all of Montana State University System’s campuses. In 1978, the library was named the Roland R. Renne Library to honor the sixth president of the university. The library supports the research and information needs of Montana’s students, faculty, and the Montana Extension Service.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia-owl:abstract</td>
<td>Montana State University Library (MSU Library) is the academic library of Montana State University, Montana's land-grant university, in Bozeman, Montana, United States. It is the flagship library for all of Montana State University System's campuses. In 1978, the library was named the Roland R. Renne Library to honor the sixth president of the university. The library supports the research and information needs of Montana’s students, faculty, and the Montana Extension Service.</td>
</tr>
</tbody>
</table>
| dbpedia-owl:wikiPageExternalLink | http://www.lib.montana.edu/  
http://www.lib.montana.edu/digital/  
http://arc.lib.montana.edu/book-0771/  
http://arc.lib.montana.edu/msu-extension/ |
| dbpedia-owl:wikiPageID    | 33974954 (xsd:integer)                                                                                                               |
| dbpedia-owl:wikiPageRevisionID | 60458055 (xsd:integer)                                                                                                           |
| dbprop:prop             | 7735532.0                                                                                                                            |
| dbprop:director         | Kenning Artisch                                                                                                                      |
| dbprop:established       | January 1894                                                                                                                          |
| dbprop:itemsCollected    | Books 523,937; E-Books 230,104; Gov Docs 11,400; Microforms-Audio-Video 2,210,828; Serial Subscription 15,615; Archivos linear |
| dbprop:libraryName       | Montana State University Library                                                                                                        |
| dbprop:location          | Bozeman, Montana, US                                                                                                                  |
| dbprop:numEmployees      | 52 (xsd:integer)                                                                                                                       |
| dbprop:phoneNum          | 994 (xsd:integer)                                                                                                                      |
| dbprop:type              | dbpedia:Academic_library                                                                                                              |
| dc:terms:subject         | category:Academic_Libraries  
category:Montana_State_University__Bozeman  
category:University_and_college_academic_libraries_in_the_United_States |
| gos:point                | 45.66670283333333 -111.04849322222222 |
| rdf:type                 | geo: SpatialThing  
owl:Thing  
http://schema.org/EducationalOrganization  
http://schema.org/Library  
http://schema.org/Organization  
http://schema.org/Place  
dd:Location  
dul:Agent  
dub:SocialPerson  
dbpedia-owl:Agent  
dbpedia-owl:ArchitecturalStructure  
dbpedia-owl:Building  
dbpedia-owl:EducationInstitution |
Remember, this was 2012
Part 3 of 3

TWO MSU SCHEMA.ORG PROJECTS
What is Schema.org?

This site provides a collection of schemas that webmasters can use to markup HTML pages in ways recognized by major search providers, and that can also be used for structured data interoperability (e.g. in JSON). Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right Web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.

We invite you to get started!

View our blog at blog.schema.org.
Schema.org

• Common vocabulary for describing things on the web
• Supported by Bing, Google, Yahoo and Yandex
• “On-page markup helps search engines understand the information on webpages and provide richer results.”

  • https://support.google.com/webmasters/answer/1211158?hl=en
Hypothesis

• Implementing Schema.org will
  – Improve machine understanding of content
  – Improve rich snippets shown in SERP
  – Increase click-through rates from SERP

  – More traffic
  – More users finding what they’re looking for
Project 1: A Controlled Experiment
by Jason Clark (with Michelle Gollehon)

• Two digital collections
• Similar size/content/date range
  – Photos and historical documents

• 1 optimized with Schema.org (Schultz)
• 1 control (Brook)
A Revised Digital Library Architecture

• Collection Page (home page)
  – arc.lib.montana.edu/schultz-0010/

• About Pages (about page, topics page)
  – arc.lib.montana.edu/schultz-0010/about.php
  – arc.lib.montana.edu/schultz-0010/topics.php

• Item Pages (individual record page)
  – arc.lib.montana.edu/schultz-0010/item/31

• Sitemap and rel=canonical work
  – arc.lib.montana.edu/schultz-0010/
About the James Willard Schultz Photos and Personal Papers Digital Collection

What is it?

James Willard Schultz (1859-1947) lived in and wrote about the northwestern portion of Montana which includes the Blackfoot Indian Reservation and Glacier National Park. In 1877, at age 18, Schultz traveled from his birthplace in Boonville, New York to Fort Benton, Montana Territory. Schultz lived for many years with the Blackfeet Indians as an accepted member of their nation, and operated several trading Posts in the Browning, Montana area. Schultz also had a guide-outfitter business in the Glacier National Park region. At age 21, Schultz started writing articles and short stories for Forest and Stream, and later authored 37 fiction and non-fiction books based upon his experiences living in Montana Territory.

The Montana State University Library’s James Willard Schultz Photograph and Personal Papers Collection includes over five hundred digital photographs of Blackfeet, Blood Indians, Kutenai, Shoshone and Arapaho Native Americans, Glacier and Waterton Lakes National Park, and historic views of Montana, Wyoming and Arizona. His personal papers include, but are not limited to: Letter and correspondence from publishers, fans, early Montana Territory friends and outfitting clients; personal journals, book drafts and editing notes, pronunciation keys for several northern-plains tribal languages, hand-drawn maps of Montana Territory, and early Newspaper Clippings.

The MSU Library James Willard Schultz Special Collections’ Finding Aid has complete details on this collection’s photographs and documents and their content and provenance. Additional James Willard Schultz Personal Papers can also be found in the University of Montana Library Archives. Montana State University (MSU) Library has also released a number of the items into the Wikimedia Commons in the "Schultz, James Willard (1859-1947) Photographs” category.

Copyright 2001-2015 Montana State University Library

These materials may be copied freely by individuals or libraries for personal use, research, teaching (including distribution to classes), or any "fair use" as defined by U.S. copyright laws. Please include this statement and author or photographer attribution with any copies you make. The materials may be linked to freely in non-commercial, non-subscription Internet editions created for an educational purpose.

Anyone interested in any other use of these materials, including for-profit Internet editions, should obtain permission from Digital Initiatives at Montana State University Library.
Waterton Lakes National Park is a national park located in the southwest corner of Alberta, Canada, and borders Glacier National Park in Montana, USA. Waterton was Canada's fourth national park, formed in 1895 and named after Waterton Lake, in turn after the Victorian naturalist and conservationist Charles Waterton. The park contains 505 km² (195 sq mi) of rugged mountains and wilderness. Operated by Parks Canada, Waterton is open all year, but the main tourist season is during July and August.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia-owl:abstract</td>
<td>Waterton Lakes National Park is a national park located in the southwest corner of Alberta, Canada, and borders Glacier National Park in Montana, USA. Waterton was Canada's fourth national park, formed in 1895 and named after Waterton Lake, in turn after the Victorian naturalist and conservationist Charles Waterton. The park contains 505 km² (195 sq mi) of rugged mountains and wilderness. Operated by Parks Canada, Waterton is open all year, but the main tourist season is during July and August.</td>
</tr>
<tr>
<td>dbpedia-owl:areaTotal</td>
<td>50500000.00000 (xsd:double)</td>
</tr>
<tr>
<td>dbpedia-owl:location</td>
<td>Alberta</td>
</tr>
<tr>
<td>dbpedia-owl:nearestCity</td>
<td>Pincher_Creek</td>
</tr>
<tr>
<td>dbpedia-owl:numberOfVisitors</td>
<td>400520 (xsd:integer)</td>
</tr>
<tr>
<td>dbpedia-owl:thumbnail</td>
<td><img src="http://commons.wikimedia.org/wiki/Special:FilePath/Watertonlake.jpg?width=300" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.pc.gc.ca/pn-np/ab/waterton/index_e.as">Http://www.pc.gc.ca/pn-np/ab/waterton/index_e.as</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.mywaterton.ca">Http://www.mywaterton.ca</a></td>
</tr>
<tr>
<td>dbpedia-owl:wikiPageID</td>
<td>720451 (xsd:integer)</td>
</tr>
<tr>
<td>dbpedia-owl:wikiPageRevisionID</td>
<td>598607441 (xsd:integer)</td>
</tr>
<tr>
<td>dbprop:areaTotalKm</td>
<td>480.580000 (xsd:double)</td>
</tr>
<tr>
<td>dbprop:coordinatesDisplay</td>
<td>inline</td>
</tr>
<tr>
<td></td>
<td>inline, title</td>
</tr>
<tr>
<td>dbprop:coordinatesRegion</td>
<td>CA-AB</td>
</tr>
<tr>
<td>dbprop:established</td>
<td>1895 (xsd:integer)</td>
</tr>
<tr>
<td></td>
<td>1979 (xsd:integer)</td>
</tr>
<tr>
<td></td>
<td>1995 (xsd:integer)</td>
</tr>
<tr>
<td>dbprop:governingBody</td>
<td>I.D. Council, Parks Canada</td>
</tr>
<tr>
<td>dbprop:hasPhotoCollection</td>
<td><a href="http://wifo5-03.informatik.uni-mannheim.de/flickr/api/photos/Waterton_Lakes_National_Park">Http://wifo5-03.informatik.uni-mannheim.de/flickr/api/photos/Waterton_Lakes_National_Park</a></td>
</tr>
<tr>
<td>dbprop:iucnCategory</td>
<td>II</td>
</tr>
<tr>
<td>dbprop:latD</td>
<td>49 (xsd:integer)</td>
</tr>
<tr>
<td>dbprop:lonM</td>
<td>2 (xsd:integer)</td>
</tr>
</tbody>
</table>
Results

Google Analytics

Sessions vs. Select a metric

Aug 1, 2014 - Oct 22, 2014: • Sessions (schultz) • Sessions (brook)
Aug 1, 2013 - Oct 22, 2013: • Sessions (schultz) • Sessions (brook)

- Wednesday, September 24, 2014
  - Sessions (schultz): 21
  - Sessions (brook): 10
- Tuesday, September 24, 2013
  - Sessions (schultz): 12
  - Sessions (brook): 7
Project 2: Schema.org IR Data Model
by Jeff Mixter and Patrick OBrien

- 1909 DC records from the Montana State University ScholarWorks IR
  - scholarworks.montana.edu
- Started with Schema.org as the base
- Created an extension vocabulary using the same mechanics and conventions used in Schema.org
  - It is published as RDFa
  - http://purl.org/montana-state/library/
Syndication of RDF data

• Loaded all 1909 RDF descriptions back into the ScholarWorks repository
  – Tweaked DSpace to pull and display JSON-ld data

• Newly created entities loaded into a Triple Store with a Pubby frontend
  – Amazon Web Services
CTR? Not so much, yet
Semantic Web Team

• Kenning Arlitsch, Dean  @kenning_msu
• Patrick OBrien, Semantic Web Director  @sempob
• Jeff Mixter, Research Associate, OCLC Research
• Jason Clark, Head of Lib Informatics and Computing @jaclark
• Scott Young, Digital Initiatives Librarian @hei_scott
• Doralyn Rossmann, Head of Coll Development @doralyn
• Jean Godby, Senior Research Associate, OCLC Research
Relevant Publications