

Libraries in Search Engines

Search Engine Optimization (SEO) for Libraries

NCompass Live
Nebraska Library Commission
September 18, 2013

Jason Clark
Head of Digital Access & Web Services
Montana State University Library



A Research Team Effort



Scott Young: Digital Initiatives Librarian at Montana State University. Scott works to make information more open and accessible through front-end web development, semantic web development, user experience research, and social media. Catch up with him on Twitter @hello_scott_.

Jason Clark: Head of Digital Access and Web Services at Montana State University Library

Patrick O'Brien: Semantic Web Research Director at Montana State University Library.

Kenning Arlitsch: Dean of the Library at Montana State University. His book, co-authored with Patrick O'Brien is titled Improving the Visibility and Use of Digital Repositories through SEO, and was published in February 2013. He occasionally tweets from @kenning_msu.

pinboard.in tag

pinboard.in/u:jasonclark/t:libseo/

twitter as channel (#hashtag)

@jaclark #libseo

Overview

- What is Search Engine Optimization (SEO)?
- Black Hat vs. White Hat SEO
- Why do SEO?
- Baseline SEO techniques
- Questions

Search Engine Optimization

- The act of creating indexable and crawlable content for commercial search engine robots

SEO



8:43 AM on July 30, 2013

A MILLION people with SEO in their job title? That fact blows me away. Great post. I'm still amazed that people are trying to game the algorithm and out-smart Google.

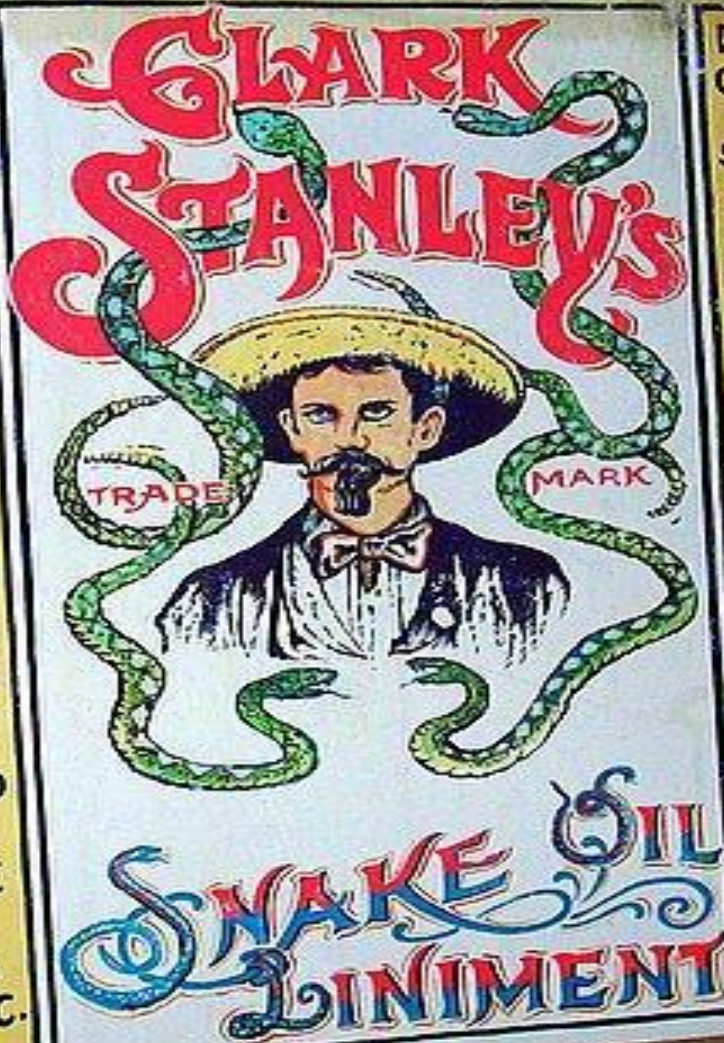
Black Hat SEO

**A
WONDERFUL
PAIN DESTROYING
COMPOUND**

THE
STRONGEST AND
BEST LINIMENT
KNOWN FOR THE
CURE OF ALL PAIN
AND LAMENESS.

USED EXTERNAL
ONLY,
FOR

RHEUMATISM
NEURALGIA
SCIATICA
LAME BACK
LUMBAGO
CONTRACTED
MUSCLES
TOOTHACHE
SPRAINS
SWELLING
ETC.



CURES
FROST BITES
CHILL BLAINS
BRUISES
SORE THROAT
BITES OF
ANIMALS
INSECTS AND
REPTILES.

IS GOOD
FOR
EVERY
THING
A
LINIMENT
SHOULD
BE GOOD
FOR.

It Gives
Immediate
Relief.

Price. 50 Cts.

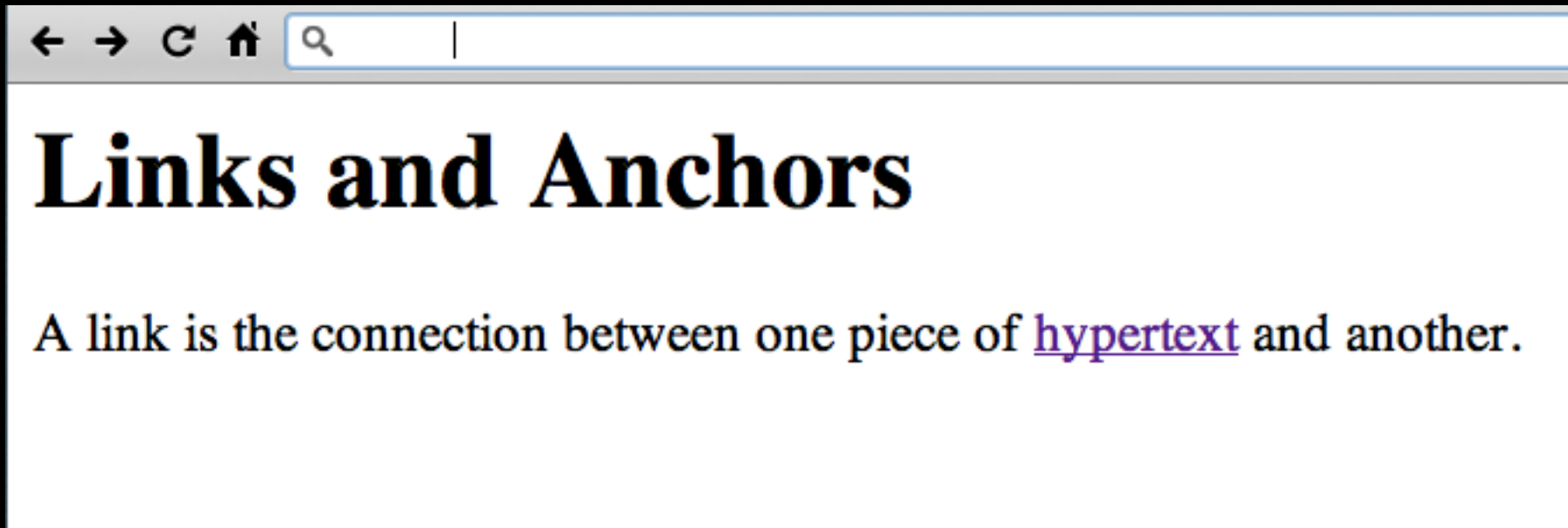
CLARK
STANLEY.
Providence,
R. I.

Black Hat SEO

- Creation of "shadow" domains that funnel users to a site by using deceptive redirects.
- "Doorway" pages loaded with keywords
- Link "farms" - pages without real content
- Hidden markup stuffed with keywords

White Hat SEO

- Approved SEO techniques
- Based on search engine guidelines and documentation for indexing
- Align with user's expectations for a search result



```
1 <title>Hypertext Links</title>
2 <h1>Links and Anchors</h1>
3 A link is the connection between one piece of
4 <a href=WhatIs.html>hypertext</a> and another.
```

Terms: HTML + CSS

Does everybody know what these elements are?

CSS

- style rules for HTML documents

HTML

- markup tags that structure docs
- browsers read them and display according to rules

"Discovery happens elsewhere."

- Lorcan Dempsey (in 2007)

Vice President and Chief Strategist of the Online Computer Library Center (OCLC)
<http://orweblog.oclc.org/archives/001430.html>

NASA



Jason Clark



+ Share

Web

Images

Maps

Shopping

News

More ▾

Search tools



About 171,000,000 results (0.44 seconds)

[NASA](#)

www.nasa.gov/ ▾

NASA.gov brings you images, videos and interactive features from the unique perspective of America's space agency. Get the latest updates on **NASA** missions, ...

[NASA Images](#)

Space Images - NASA Photography -
Mars Images - ...

[NASA Television](#)

NASA Television provides live
coverage of launches ...

[NASA News and Features](#)

Get NASA Breaking News ...
NEWSNews, features & press ...

[International Space Station](#)

Orbital Tracking - Video - Expedition
36 - Karen L. Nyberg

[NASA Video Gallery](#)

Skip to main content. LOADING...
LOADING... NEWSNews ...

[NASA Kids](#)

Buzz Lightyear - Mars - Elmo Visits
NASA - Text Only Site - ...

Search nasa.gov

NASA

667,526 followers on Google+

Follow



The National Aeronautics and Space Administration is the agency of the United States government that is responsible for the nation's civilian space program and for aeronautics and aerospace research. [Wikipedia](#)

Founded: July 29, 1958

Founders: [Dwight D. Eisenhower](#), [John F. Kennedy](#), [Wernher von Braun](#), [Richard Nixon](#)



Scholar

11 results (0.07 sec)

Articles

Legal documents

Any time

Since 2013

Since 2012

Since 2009

Custom range...

Sort by relevance

Sort by date

☒ include patents

☒ include citations

☐ Create alert

[Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar](#)

[PDF] from duraspac

[K Arlitsch](#), [PS O'Brien](#) - Library Hi Tech, 2012 - emeraldinsight.com

Purpose—Google Scholar has difficulty indexing the contents of **institutional repositories**, and the authors hypothesize the reason is that most **repositories** use Dublin Core, which cannot express bibliographic citation information adequately for academic papers. Google ...

Cited by 16 Related articles All 6 versions Cite

[Aggregating distributed digital collections in the Mountain West Digital Library with the CONTENTdmTM multi-site server](#)

[K Arlitsch](#), [J Jonsson](#) - Library Hi Tech, 2005 - emeraldinsight.com

... This lightweight data harvesting method is, in effect, **invisible** to the remote sites ... Cooperation and a respect for the differences in participants' missions, **institutional** cultures, and funding structures are ... Of the 30 GWLA **institutions**, 12 are actively participating in the two-year project ...

Cited by 13 Related articles All 5 versions Cite

[Automating the production of map interfaces for digital collections using Google APIs](#)

[HTML] from uniriioja

[A Neatrou](#), [A Morrow](#), [K Rockwell](#), [A Witkowski](#) - D-Lib magazine, 2011 - dialnet.uniriioja.es

... While the second step in the process applies directly to **institutions** using CONTENTdm as the ... State Historical Society and a map based on the content of the **Hidden** Water collection ... Anne has a background in digital libraries, **institutional repositories**, metadata and cataloging. ...

Cited by 3 Related articles All 11 versions Cite More▼

[Success factors and strategic planning: Rebuilding an academic library digitization program](#)

[PDF] from unlv.edu

[CK Lampert](#), [J Vaughan](#) - Information Technology and ..., 2009 - digitalscholarship.unlv.edu

... For **institutions** with recognized digitization staff, great variations exist between **institutions** in terms of ... the biggest factor for success of a digital library project." **Institutional** priorities at ... are uncataloged, and digitizing such collections would help promote these **hidden** treasures ...

"Search engines continue to dominate, topping the list of electronic sources most used to find online content (93%), followed closely by Wikipedia (88%). The key difference in usage between search engines and Wikipedia is the frequency - 75% of students who use search engines do so daily, compared to 20% of those who use Wikipedia."

Perceptions of Libraries, 2010: Context and Community: a Report to the OCLC Membership. OCLC, 2011.

http://www.oclc.org/content/dam/oclc/reports/2010perceptions/2010perceptions_all_singlepage.pdf

<http://www.oclc.org/content/dam/oclc/reports/2010perceptions/collegestudents.pdf>

SEO - Baseline Techniques

Keyword analysis

Writing titles and descriptions

Creating indexable content

- sitemaps, linked architecture

Reduce noise in the index

- rel canonical, noindex nofollow

Introduce semantic markup

Keyword Analysis

- Research your market
- Write or rewrite your web copy, titles, descriptions based on this research
- Set up Google Adwords Account
 - Keyword Planner

Writing Titles and Descriptions

Writing web copy for the robot and the human

```
<title></title>
```

```
<meta name="description" />
```

Writing Titles and Descriptions

Title

"Keyword Phrase | Context"

Example:

"Staff Directory and People Listing: Montana State University MSU Library"

Writing Titles and Descriptions

Description

a short, declarative sentence that incorporates your keywords, as you laid it out in your page title

Example:

"A listing of staff employed at Montana State University MSU Library including library departments, roles, job titles, phone numbers and contact information."

Creating Indexable Content

- Clear hierarchy and site architecture
 - Directory in URL shows hierarchy
 - Breadcrumb links

“Every page should be reachable from at least one static text link.”

<https://support.google.com/webmasters/answer/35769>

Creating Indexable Content

- Set up machine-readable sitemap
 - Listing all pages that you want to be indexed and indexing priorities for those pages

“Google adheres to Sitemap Protocol 0.9 as defined by sitemaps.org.”

<https://support.google.com/webmasters/answer/156184>

Creating Indexable Content

- Set up machine-readable sitemap
 - Listing all pages that you want to be indexed and indexing priorities for those pages

“Google adheres to Sitemap Protocol 0.9 as defined by sitemaps.org.”

<https://support.google.com/webmasters/answer/156184>

Reduce “Noise” in the Index

- Redundant content
- Dynamic pages, search result pages
- Not every page is unique

Techniques:

1. robots.txt directives
2. link relations and index directive markup

Reduce “Noise” in the Index

robots.txt

- set of commands in text file at top level of your site

```
Disallow: /staff/
```

```
#allow primary mobile page
```

```
Allow: /finding-aids/m/index.php
```

```
Allow: /finding-aids/m/$
```

Reduce “Noise” in the Index

Link relations and index directive markup

- markup that identifies canonicalization or whether to index a page

```
<link rel="canonical"
href="http://arc.lib.montana.edu/finding-
aids/item/23" />
```

```
<a rel="nofollow"
href="http://arc.lib.montana.edu/finding-
aids/item/23"></a>
```

```
<meta name="robots" content="nofollow,noindex" />
```

Introducing Semantic Markup

- HTML5 semantic tags and microdata that helps classify page types and types of content on the page

“If Google understands the content on your pages, we can create rich snippets—detailed information intended to help users with specific queries.”

https://support.google.com/webmasters/answer/99170?hl=en&ref_topic=1088472

Item: **Women posing on college locomotive engine****Title:** Women posing on college locomotive engine**Creator:** unknown**Date:** unknown**Description:** Women posing on a college locomotive engine.**Notes:****Physical Description:** Photo print - Black and White**Subjects:** [Locomotive](#)**Keywords:** locomotive**Photograph ID:** parc-000432 [Higher quality/resolution](#) [Persistent link](#) [Bookmark and Share](#) [Get embed code](#)[<a title="Women posing on college locomo](#)[+ Back to MSU Photos Homepage](#)

BM (Before Microdata)

Item: **Women posing on college locomotive engine****Title:** Women posing on college locomotive engine**Creator:** unknown**Date:** unknown**Description:** Women posing on a college locomotive engine.**Notes:****Physical Description:** Photo print - Black and White**Subjects:** [Locomotive](#)**Keywords:** locomotive**Photograph ID:** parc-000432 [Higher quality/resolution](#) [Persistent link](#) [Bookmark and Share](#) [Get embed code](#)[<a title="Women posing on college loco](#)[+ Back to MSU Photos Homepage](#)

AM (After Microdata)

```
<dl>
<dt>Title</dt>
<dd>A River Runs Through It and Other
Stories</dd>
<dt>Author</dt>
<dd>Norman Maclean</dd>
<dt>Publication date</dt>
<dd>October 1, 2001</dd>
<dt>ID</dt>
<dd>0226500667</dd>
</dl>
```



```
<dl itemscope itemtype="http://schema.org/Book">
<dt>Title</dt>
<dd itemprop="title">A River Runs Through It and
Other Stories</dd>
<dt>Author</dt>
<dd itemprop="author">Norman Maclean</dd>
<dt>Publication date</dt>
<dd itemprop="pubdate">October 1, 2001</dd>
<dt>ID</dt>
<dd itemprop="isbn">0226500667</dd>
</dl>
```

Schema.org

A lingua franca for web classification and markup

Google, Bing, Yahoo

SEO in Libraries

Arlitsch, Kenning, and Patrick S. O'Brien. "Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar." *Library Hi Tech* 30.1 (2012): 60-81.

Onaifo, Daniel, and Diane Rasmussen. "Increasing libraries' content findability on the web with search engine optimization." *Library Hi Tech* 31.1 (2013): 87-108.

Improving the Visibility and Use of Digital Repositories Through Seo



Kenning Arlitsch, Patrick S. O'Brien

★★★★★

0 Reviews

NEAL-SCHUMAN PUB Incorporated, 2013 - [Computers](#) - 128 pages

Recent OCLC surveys show that less than 2 percent of library users start their search on a library website. Another survey of faculty researchers at four major universities showed that most consider Google and Google Scholar amazingly effective for their research. Low Google Scholar indexing ratios for library institutional repositories is widespread because it ignores common library metadata, and high-value content through libraries is consequently invisible to researchers. Authors Arlitsch and O'Brien share their expertise in digital libraries and corporate marketing to offer practical steps for search engine optimization, such as: Recommended dashboards to increase participation by sharing data Avoiding the four most common crawler errors that lead to low rankings How to effectively utilize the Google Keyword Tool How to use domain settings to generate unit-specific reports for special collections, institutional repositories, and university presses

Arlitsch, Kenning, and Patrick S. O'Brien. Improving the Visibility and Use of Digital Repositories Through Seo: A Lita Guide. Chicago: ALA TechSource, 2013. Print.

Questions?

twitter.com/jaclark

www.lib.montana.edu/~jason/talks.php