RDFa Markup, Schema.org Vocabularies, and DBpedia Topics for Digital Collections

A Closer Look at the Holy Trinity of Structured Data and their Impact on the Findability of Digital Collections

"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.

Website's Built-In Search Engine Just Pathetic

NEWS IN BRIEF · Science & Technology · Internet · ISSUE 49·30 · Jul 22, 2013

CHICAGO—Following several unsuccessful attempts to find a specific article on a national media website, readers confirmed Monday that the built-in search engine found in the upper-right-hand corner of the online publication is nothing short of useless. "For example, I wanted to find a story they did recently about Michael Bay, and so I typed in his name, but the first results that came up were articles about NASA, Jim Harbaugh, and aliens," said longtime reader Courtney Dunning, who admitted to reporters she was surprised that a website with two decades of backlogged content would have such a poorly designed and ineffective method for accessing its archived material. "Searching for 'Michael Bay's new movie' didn't get me anywhere, either. In fact, even when I typed in the exact headline of the article, it told me no results could be found." As of press time, Dunning had reportedly abandoned the website's search bar, Googled what she was looking for, and found it on her first attempt. ⌀

As of press time, Dunning had reportedly abandoned the website's search bar, Googled what she was looking for, and found it on her first attempt. 🌱

The Project

Designing Your Digital Collections for Web-Scale Search

Designing Your Library Collections for Web-Scale Search

Modeling best practices for digital collections markup

Creating metadata routines for web-scale external vocabularies

Measuring the impact of applying external vocabularies to digital collections

A Controlled Experiment

- Two collections - 1 optimized, 1 left on its own
- Similar content - photos & historical documents
- Similar size - Around 800 to 1000 items
- Similar date range - Mid 19th to Mid 20th century
- Similar topical location - Montana

arc.lib.montana.edu/schultz-0010
arc.lib.montana.edu/brook-0771

```html
<!doctype html>
<html lang="en" vocab="http://schema.org/" typeof="CreativeWork">
<head vocab="http://schema.org/" typeof="WebPage">
<meta charset="utf-8">
<meta http-equiv="X-UA-Compatible" content="IE=edge">
<meta name="viewport" content="width=device-width, initial-scale=1">
<title property="name">Topics for The James Willard Schultz Photos and Personal Papers Collection - Montana State University (MSU) Library</title>
<meta name="description" property="description" content="List of topics in the James Willard Schultz Photos and Personal Papers Digital Collection in the Montana
State University Library Archives." />
<link rel="canonical" href="http://arc.lib.montana.edu/schultz-0010/topics.php" />
<link rel="stylesheet" type="text/css" media="screen, projection" href="/schultz-0010/meta/styles/master.css" />
</head>
<body class="fullWidth default">
<div id="contain">
<div id="mastHead">
  <h1><a href="/schultz-0010/index.php">James Willard Schultz Photos and Personal Papers - Montana State University (MSU) Library</a></h1>
  <p class="skipNav"><a href="#mainContent" title="main content">skip navigation</a></p>
  <p><a href="http://www.lib.montana.edu/digital/">MSU Library Online Collections</a> | <a href="/schultz-0010/m/">Mobile</a></p>
</div><!-- end masthead div -->
<div id="main" vocab="http://schema.org/" typeof="AboutPage">
<meta property="isPartOf" content="http://arc.lib.montana.edu/schultz-0010/" />
<a id="mainContent"></a>
<div class="gutter">
<h2 class="mainHeading"><span property="name">Topics for The James Willard Schultz Photos and Personal Papers Collection</span></h2>
  <ul id="listColumns" property="mainContentOfPage">
      <li property="additionalType" typeof="URL" resource="http://dbpedia.org/page/Alberta">
        <a href="./search.php?topic=%22Alberta%22"><span property="specialty">Alberta</span> (7)</a>
        <link property="url" href="http://dbpedia.org/page/Alberta" />
      </li>
      <li property="additionalType" typeof="URL" resource="http://dbpedia.org/page/American_Bison">
        <a href="./search.php?topic=%22American+Bison%22"><span property="specialty">American Bison</span> (6)</a>
        <link property="url" href="http://dbpedia.org/page/American_Bison" />
      </li>
...
</div><!-- end main div -->
<div id="footer">
<ul>
  <li id="info">&copy; Copyright 2001-2014 <span property="copyrightHolder publisher">Montana State University (MSU) Library</span></li>
  <li id="links"><a class="contact" href="http://www.lib.montana.edu/digital/contact.php" title="link to contact us">Contact Us</a> | <a class="xml" href="/
  schultz-0010/feed.xml" title="Subscribe to Topics for The James Willard Schultz Photos and Personal Papers Collection feed" rel="alternate" type="application/
  atom+xml">XML feed</a></li>
</ul>
</div><!-- end footer div -->
</div><!-- end contain div -->
</body>
</html>
```

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.

DBpedia

Le Seigneur des anneaux (The Lord of the Rings) est un roman en trois volumes de J. R. R. Tolkien paru en 1954 et 1955.L'histoire reprend certains des personnages présentés dans Le Hobbit, premier roman de l'auteur paru en 1937, mais l'œuvre est plus complexe et plus sombre. Tolkien entreprend sa rédaction à la demande de son éditeur, Allen & Unwin, à la suite du succès critique et commercial du Hobbit.

Property	Value
dbpedia-owl:abstract	- The Lord of the Rings is an epic high fantasy novel written by English author J. R. R. Tolkien. The story began as a sequel to Tolkien's 1937 children's fantasy novel The Hobbit, but eventually developed into a much larger work. It was written in stages between 1937 and 1949, much of it during World War II. It is the second best-selling novel ever written, with over 150 million copies sold.The title of the novel refers to the story's main antagonist, the Dark Lord Sauron, who had in an earlier age created the One Ring to rule the other Rings of Power as the ultimate weapon in his campaign to conquer and rule all of Middle-earth. From quiet beginnings in the Shire, a hobbit land not unlike the English countryside, the story ranges across northwest Middle-earth, following the course of the War of the Ring through the eyes of its characters, the hobbits Frodo Baggins, Samwise "Sam" Gamgee, Meriadoc "Merry" Brandybuck and Peregrin "Pippin" Took, but also the hobbits' chief allies and travelling companions: the Men Aragorn, a Ranger of the North and Boromir, a Captain of Gondor; Gimli, a Dwarf warrior; Legolas, an Elven prince; and Gandalf, a Wizard.The work was initially intended by Tolkien to be one volume of a two-volume set, the other to be The Silmarillion, but this idea was dismissed by his publisher. For economic reasons The Lord of the Rings was published in three volumes over the course of a year from 29 July 1954 to 20 October 1955. The three volumes were titled The Fellowship of the Ring, The Two Towers, and The Return of the King. Structurally, the novel is divided internally into six books, two per volume, with several appendices of background material included at the end of the third volume. Some editions combine the entire work into a single volume. The Lord of the Rings has since been reprinted numerous times and translated into many languages.Tolkien's work has been the subject of extensive analysis of its themes and origins. Although a major work in itself, the story was only the last movement of a larger epic Tolkien had worked on since 1917, in a process he described as mythopoeia.[citation needed] Influences on this earlier work, and on the story of The Lord of the Rings, include philology, mythology, religion and the author's distaste for the effects of industrialization, as well as earlier fantasy works and Tolkien's experiences in World War I. The Lord of the Rings in its turn is considered to have had a great effect on modern fantasy; the impact of Tolkien's works is such that the use of the words "Tolkienian" and "Tolkienesque" have been recorded in the Oxford English Dictionary.The enduring popularity of The Lord of the Rings has led to numerous references in popular culture, the founding of many societies by fans of Tolkien's works, and the publication of many books about Tolkien and his works. The Lord of the Rings has inspired, and continues to inspire, artwork, music, films and television, video games, and subsequent literature. Award-winning adaptations of The Lord of the Rings have been made for radio, theatre, and film. - Le Seigneur des anneaux (The Lord of the Rings) est un roman en trois volumes de J. R. R. Tolkien paru en 1954 et 1955.L'histoire reprend certains des personnages présentés dans Le Hobbit, premier roman de l'auteur paru en 1937, mais l'œuvre est plus complexe et plus sombre. Tolkien entreprend sa rédaction à la demande de son éditeur, Allen & Unwin, à la suite du succès critique et commercial du Hobbit. Il lui faut douze ans pour parvenir à achever ce nouveau roman qu'il truffe de références et d'allusions au monde du Silmarillion, la Terre du Milieu, sur lequel il travaille depuis 1917 et dans lequel Le Hobbit a été attiré de son auteur.C'est une des œuvres fondamentales de la littérature dite de fantasy, terme que Tolkien explicite dans son essai Du conte de fées de 1939. Tolkien lui-même considérait son livre comme , écrit .Cette œuvre est composée de six livres, qui ne portent pas de titres. À l'origine, Tolkien souhaite publier Le Seigneur des anneaux en un seul volume, mais le prix du papier étant trop élevé en cette période d'après-guerre, l'œuvre est divisée en trois volumes : La Communauté de l'anneau (The Fellowship of the Ring), Les Deux Tours (The Two Towers) et Le Retour du roi (The Return of the King). On fait souvent référence à cette œuvre comme à « la trilogie du Seigneur des anneaux », terme techniquement incorrect car l'œuvre fut écrite et conçue d'un seul tenant. Néanmoins, Tolkien lui-même reprend dans ses lettres, de temps à autres, le terme de « trilogie » lorsqu'il est employé par ses correspondants.
dbpedia-owl:wikiPageExternalLink	- http://www.houghtonmifflinbooks.com/features/lordoftheringstrilogy/ - http://www.glyphweb.com/arda/ - http://www.tolkien.co.uk - http://www.tolkienlibrary.com
dbpedia-owl:wikiPageID	• 29798 (xsd:integer)
dbpedia-owl:wikiPageRevisionID	• 606679234 (xsd:integer)
dbpprop:author	• dbpedia:J._R._R._Tolkien
dbpprop:books	• *Volumes: *The Fellowship of the Ring *The Two Towers *The Return of the King
dbpprop:country	• United Kingdom
dbpprop:genre	• *High fantasy *Adventure
dbpprop:hasPhotoCollection	• http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/photos/The_Lord_of_the_Rings
dbpprop:imageCaption	• The original cover designs for each volume as illustrated by Tolkien. They were later used for the 50th anniversary edition covers.
dbpprop:language	• English
dbpprop:mediaType	• Print
dbpprop:name	• The Lord of the Rings

Item: **Letter to James Willard Schultz about the health of George Bird Grinnell**

Title: Letter to James Willard Schultz about the health of George Bird Grinnell

Creator: George E. Sheetz

Date: 1937-02-23

Description: A letter written by George E. Sheetz, secretary to Mr. George Bird Grinnell, to Mr. James Willard Schultz in response to his many letters. An explanation of Mr. Grinnell's health is given and a request to receive the stories Schultz wrote for the Great Falls Tribune.

Notes: None

Object's Physical Description: application/pdf

Keywords: james willard schultz, letter, newspaper

Topic(s): George Bird Grinnell

Related Link
https://en.wikipedia.org/wiki/George_Bird_Grinnell

Object ID: 0010-B06-F15-79

Document PDF
Persistent link
Bookmark and share
Embed as HTML
<a title="Letter to James Willard Schultz about the he

← Back to Schultz Collection Homepage

Wait, what… how?

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/

External Enumerations

"We define here some specific integration points through which selected externally maintained vocabulary can be published as part of schema.org markup"

www.w3.org/wiki/WebSchemas/ExternalEnumerations

Controlled vocabulary

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(June 2012)*

Controlled vocabularies provide a way to organize knowledge for subsequent retrieval. They are used in subject indexing schemes, subject headings, thesauri, taxonomies and other forms of knowledge organization systems. Controlled vocabulary schemes mandate the use of predefined, authorised terms that have been preselected by the designer of the vocabulary, in contrast to natural language vocabularies, where there is no restriction on the vocabulary.

Contents [hide]

In library and information science [edit]

In library and information science controlled vocabulary is a carefully selected list of words and phrases, which are used to tag units of information (document or work) so that they may be more easily retrieved by a search.[1][2] Controlled vocabularies solve the problems of homographs, synonyms and polysemes by a bijection between concepts and authorized terms. In short, controlled vocabularies reduce ambiguity inherent in normal human languages where the same concept can be given different names and ensure consistency.

Wikipedia (DBpedia) as a Controlled Vocabulary

github.com/jasonclark/getsemantic

additionalTypes, specialty

```
<li property="additionalType" typeof="URL" resource="http://dbpedia.org/page/Sacagawea">
  <a href="./search.php?topic=Sacagawea"><span property="specialty">Sacagawea</span></a>
  <link property="url" href="http://dbpedia.org/page/Sacagawea" />
</li>
```

https://schema.org/additionalType, http://schema.org/specialty
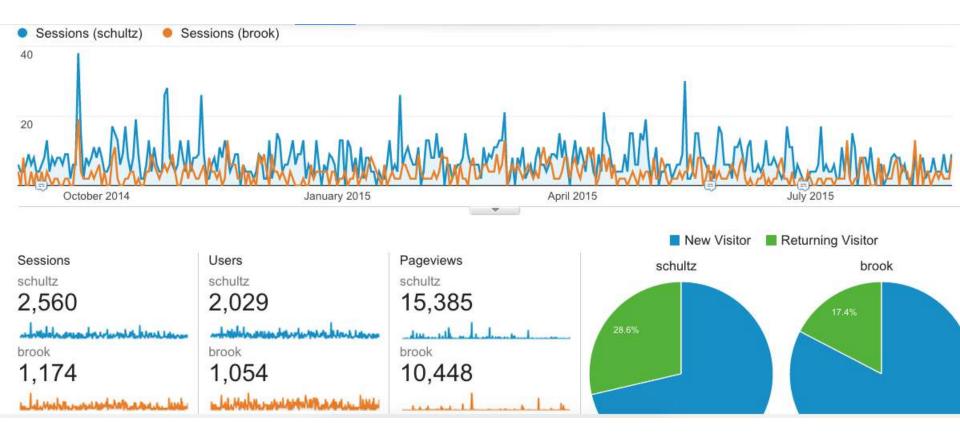
CollectionPage, AboutPage, ItemPage

```
Index   page
<link rel="canonical" href="http://arc.lib.montana.edu/schultz-0010/" />
...
<div id="main" vocab="http://schema.org/" typeof="CollectionPage">
```

```
Topics page
<div id="main" vocab="http://schema.org/" typeof="AboutPage">
<meta property="isPartOf" content="http://arc.lib.montana.edu/schultz-0010/" />
```

```
Item page
<div id="main" vocab="http://schema.org/" typeof="ItemPage">
<meta property="isPartOf" content="http://arc.lib.montana.edu/schultz-0010/" />
```

https://schema.org/CollectionPage, http://schema.org/AboutPage, https://schema.org/ItemPage

Results

http://arc.lib.montana.edu/schultz-0010.

FETCH & VALIDATE   CANCEL   Shortlink

Results - Filter by use case ▾

```
30  <meta property="thumbnailUrl"
    content="http://arc.lib.montana.edu/schultz-0010/objects/0010-B06-
    F15-79.jpg" />
31  <div id="main" vocab="http://schema.org/" typeof="ItemPage">
32  <meta property="isPartOf"
    content="http://arc.lib.montana.edu/schultz-0010/schultz-0010/" />
33      <a id="mainContent"></a>
34      <div class="gutter" property="mainContentOfPage">
35              <h2 class="mainHeading">Item: <strong><span
    property="name">Letter to James Willard Schultz about the health of
    George Bird Grinnell</span></strong></h2>
36              <ul class="item" vocab="http://schema.org/"
    typeof="MediaObject">
37                      <li>
38                          <ul class="metadata">
39                              <li class="object">
40                              <img property="image" src="/schultz-
    0010/objects/0010-B06-F15-79.jpg" alt="Letter to James Willard
```

▾ ItemPage (1)   All good ✓

ItemPage

name: Letter to James Willard Schultz about the health of George B
                 ird Grinnell

isPartOf [CreativeWork]:

  name: http://arc.lib.montana.edu/schultz-0010/schultz-0010/

mainContentOfPage [WebPageElement]:

  name: Item: Letter to James Willard Schultz about the health of
                 George Bird Grinnell Title: Letter to James Willard Schul
                 tz about the health of George Bird Grinnell Creator: Geo
                 rge E. Sheetz Date: 1937-02-23 Description: A letter wri
                 tten by George E. Sheetz, secretary to Mr. George Bird
                 Grinnell, to Mr. James Willard Schultz in response to his
                 many letters. An explanation of Mr. Grinnell's health is g
                 iven and a request to receive the stories Schultz wrote f

Home

Structured Data Testing Tool

Structured Data Markup Helper

Help with:

Troubleshooting

Help Center

Structured Data Testing Tool

URL	HTML

http://arc.lib.montana.edu/schultz-0010/item/702

PREVIEW

Examples ▾

Google search results	Google Custom Search

Extracted structured data

rdfa-node

type: ItemPage

property:

isPartOf: http://arc.lib.montana.edu/schultz-0010/schultz-0010/

mainContentOfPage: Item: Letter to James Willard Schultz about the health of George Bird Grinnell Title: Letter to James Willard Schultz about the health of George Bird Grinnell Creator: George E. Sheetz Date:...

name: Letter to James Willard Schultz about the health of George Bird Grinnell

rdfa-node

type: MediaObject

relationship:

name: image

value: Letter to James Willard Schultz about the health of George Bird Grinnell

Webmaster Tools

Site Dashboard

Site Messages

▸ Search Appearance ⓘ

▸ Search Traffic

▾ **Google Index**

   Index Status

   Content Keywords

   Remove URLs

▸ Crawl

Security Issues

Other Resources

Content Keywords

Keyword	Significance
1. schultz (2 variants)	
2. library (3 variants)	
3. msu	
4. montana	
5. photos (2 variants)	
6. collection (2 variants)	
7. state (3 variants)	
8. university	
9. james	
10. willard	

	A	B	C
25	glacier	193	glacier, glaciers
26	black	191	black
27	bookmark	184	bookmark
28	blackfeet	163	blackfeet
29	tipi	156	tipi, tipis
30	lodge	144	lodge, lodges
31	standing	143	standing, stand, stands
32	women	143	women
33	browning	137	browning
34	copyright	135	copyright
35	man	132	man
36	front	126	front
37	america	125	america
38	letter	118	letter
39	code	110	code
40	mountains	106	mountains, mountain
41	men	100	men

Machine Understanding
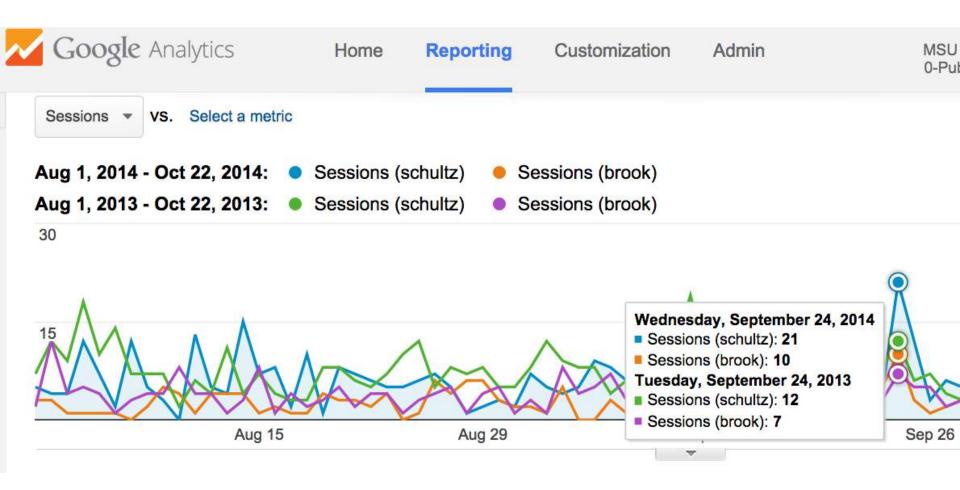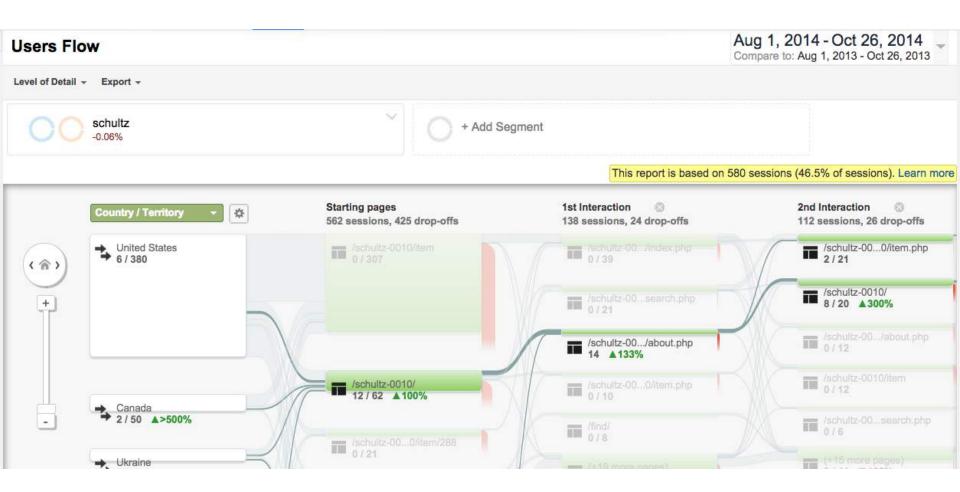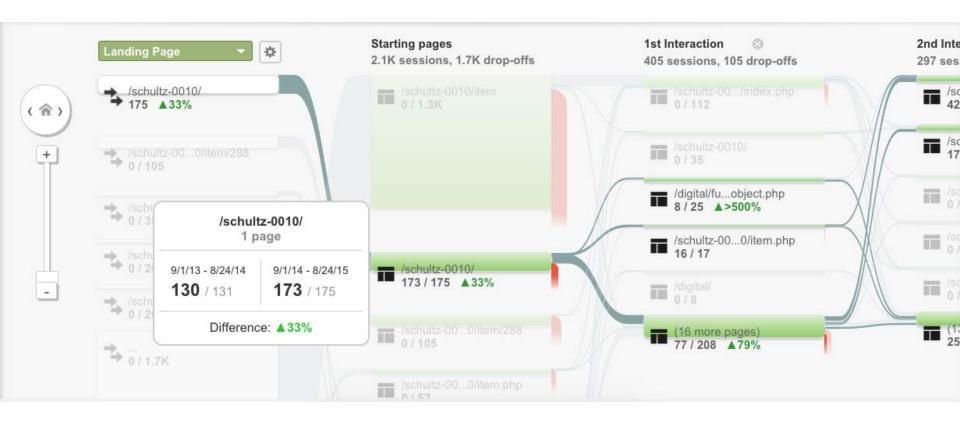
Snapshot of Search Behavior (May 26 to August 23, 2015)

Room for improvement based on last slide…

impressions are solid (appearance in search result)
clicks and CTR are average/low (engagement)

Level of Detail ▾   Export ▾

schultz
-0.06%

+ Add Segment

This report is based on 580 sessions (46.5% of sessions). Learn more

Country / Territory ▾   ⚙

Starting pages
562 sessions, 425 drop-offs

1st Interaction ⊗
138 sessions, 24 drop-offs

2nd Interaction ⊗
112 sessions, 26 drop-offs

United States
6 / 380

/schultz-0010/item
0 / 307

/schultz-00.../index.php
0 / 39

/schultz-00...0/item.php
2 / 21

/schultz-00...search.php
0 / 21

/schultz-0010/
8 / 20 ▲300%

/schultz-00.../about.php
14 ▲133%

/schultz-00.../about.php
0 / 12

/schultz-0010/
12 / 62 ▲100%

/schultz-00...0/item.php
0 / 10

/schultz-0010/item
0 / 12

Canada
2 / 50 ▲>500%

/find/
0 / 8

/schultz-00...search.php
0 / 6

/schultz-00...0/item/288
0 / 21

(+19 more pages)

(+15 more pages)

Ukraine

Future Goals

Set up RDFa model so that each item page can be converted to linked data triples

Streamline metadata routines and apply model to other library collections

Use longitudinal analytics data to measure impact, make improvements to model, and assess utility of the approach

Sara Mannheimer - People at MSU Library

Sara Mannheimer

Title Data Management Librarian [Assistant Professor]

Department Library Informatics and Computing, Montana State University (MSU)

Phone 406-994-3361

Email sara.mannheimer@montana.edu

Room 135

Liaison Role(s)

Computer science

Expertise and Skill(s)

Data Management
Institutional Repository
Open Access
Digital Preservation

More Information

Graph Visualization
Vita
RDF

← Back to Staff Home page

Eva Méndez
@evamen

☼   +&   Follow

OpenGraph and Schema.org is a déjà vu...
do you remember Altavista <meta name
="keywords">? nope... you are too young
#dcmi14

↩   ↻   ★   •••

RETWEETS   FAVORITES
2   6

3:46 PM - 11 Oct 2014

AYYY!

Questions

Michelle Gollehon

Library Tech at Montana State University

Bozeman, Montana (Great Falls, Montana Area) | Museums and Institutions

- Controlled Vocabulary Work
- Sounding Board

Semantic Web Group

Scott Young: Digital Initiatives Librarian at Montana State University Library.
@hei_scott

Jason Clark: Head of Library Informatics & Computing at Montana State University Library
@jaclark

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

Kenning Arlitsch: Dean of the Library at Montana State University
@kenning_msu

Doralyn Rossmann: Head of Collection Development at Montana State University Library
@doralyn

Resources for Digital Collection Markup + Testing

- https://developers.google.com/structured-data/testing-tool/
- https://schema.org/AboutPage
- https://schema.org/CollectionPage
- https://schema.org/ItemPage
- http://www.w3.org/wiki/WebSchemas/additionalTypeProposal
- http://www.w3.org/wiki/WebSchemas/MultipleTypesSDO
- http://www.w3.org/wiki/WebSchemas/ExternalEnumerations

Why do SEO?

- ~200 million people in the US use the Internet[1]
  - 81% of U.S. adults
  - 95% of U.S. teens
- The US submits 19+ Billion queries per month to commercial SE[2]
- Google owns 67% of these queries

[1] Pew Research Center. "What Internet users do online: Pew Research Center's Internet & American Life Project," 2012. http://pewinternet.org/Trend-Data-(Adults)/Online-Activites-Total.aspx

[2] comScore. "comScore Releases February 2014 U.S. Search Engine Rankings," February 18, 2014. http://www.comscore.com/Insights/Press_Releases/2014/2/comScore_Releases_January_2014_US_Search_Engine_Rankings