Findability in the Flow: Discovery through Linking

Mark A. Matienzo
Manuscripts and Archives, Yale University Library
Society of American Archivists Annual Meeting
Session 701 — August 14, 2010
mark@matienzo.org
http://matienzo.org/
twitter: @anarchivist
Disclaimer

The following presentation expresses opinions of my own and not of my employer, my coworkers, etc.
Why are you in this session?
Why are you interested in SEO?

http://www.flickr.com/photos/keithmarshall/1302601776/
This isn't the only goal
You want to be found.

http://www.flickr.com/photos/lakelandlocal2/3204038379
Search or discovery?

http://www.flickr.com/photos/wwwworks/1431384410
<table>
<thead>
<tr>
<th>Search</th>
<th>Discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Implies seeking a known object</td>
<td>• Implies exploration for underspecified or</td>
</tr>
<tr>
<td></td>
<td>unknown objects</td>
</tr>
<tr>
<td>• Seeker has specific characteristics in</td>
<td>• General characteristics outline information space</td>
</tr>
<tr>
<td>mind</td>
<td></td>
</tr>
<tr>
<td>• Characteristics are used to specify and</td>
<td>• Info. space examined w/ perception/cognition</td>
</tr>
<tr>
<td>guide search activity</td>
<td></td>
</tr>
<tr>
<td>• Analytical behavior</td>
<td>• Browsing behavior</td>
</tr>
</tbody>
</table>

Discovery happens elsewhere

Lorcan Dempsey

http://orweblog.oclc.org/archives/001430.html

http://www.flickr.com/photos/nengard/4399134103/
Go where the users are!

http://www.flickr.com/photos/seeminglee/3921229872/
URLs are the currency of the web.

Lorcan Dempsey
http://orweblog.oclc.org/archives/000984.html

http://www.flickr.com/photos/penguincakes/2930270559/
Users exchange links for information
Users exchange links for influence
Users exchange links for goodwill

http://www.flickr.com/photos/ryanr/142454711
Links are important for machines, too
Web crawlers can help define relationships
Web crawlers can help give new insights
Web crawlers can help build reputation
Look for backlinks into your resources

http://www.flickr.com/photos/snapeverything/2475894906/
Follow your nose to other resources
Examples
Like Button

The Like button lets a user share your content with friends on Facebook. When the user clicks the Like button on your site, a story appears in the user's friends' News Feed with a link back to your website.

When your Web page represents a real-world entity, things like movies, sports teams, celebrities, and restaurants, use the Open Graph protocol to specify information about the entity. If you include Open Graph tags on your Web page, your page becomes equivalent to a Facebook page. This means when a user clicks a Like button on your page, a connection is made between your page and the user. Your page will appear in the "Likes and Interests" section of the user's profile, and you have the ability to publish updates to the user. Your page will show up in same places that Facebook pages show up around the site (e.g. search), and you can target ads to people who like your content.

There are two Like button implementations: XFBML and Iframe. The XFBML version is more versatile, but requires use of the JavaScript SDK. The XFBML dynamically re-sizes its height according to whether there are profile pictures to display, gives you the ability (through the Javascript library) to listen for like events so that you know in real time when a user clicks the Like button, and it always gives the user the ability to add an optional comment to the like. If users do add a comment, the story published back to Facebook is given more prominence.

To get started, just use the configurator below to get code to add to your site.

http://developers.facebook.com/docs/reference/plugins/like
You can “like” real things
Open Graph protocol

Introduction

The Open Graph protocol enables you to integrate your Web pages into the social graph. It is currently designed for Web pages representing profiles of real-world things — things like movies, sports teams, celebrities, and restaurants. Once your pages become objects in the graph, users can establish connections to your pages as they do with Facebook Pages. Based on the structured data you provide via the Open Graph protocol, your pages show up richly across Facebook: in user profiles, within search results and in News Feed.

Getting Started

To turn your web pages into graph objects, you'll need to add Open Graph protocol <meta> tags and the Like button to your webpages.

The tags allow you to specify structured information about your web pages. The more information you provide, the more opportunities your web pages can be surfaced within Facebook today and in the future. Here's an example for a movie page:

```
<html xmlns:og="http://opengraphprotocol.org/schema/
xmlns:fb="http://www.facebook.com/2008/fbml">
<head>
<title>The Rock (1996)" /></title>
<meta property="og:title" content="The Rock" />
```

http://developers.facebook.com/docs/opengraph
<html xmlns:og="http://opengraphprotocol.org/schema/"
     xmlns:fb="http://www.facebook.com/2008/fbml">
  <head>
    <title>The Rock (1996)</title>
    <meta property="og:title" content="The Rock"/>
    <meta property="og:type" content="movie"/>
    <meta property="og:url" content="http://www.imdb.com/title/tt0117500/"/>
    <meta property="og:image" content="http://ia.media-imdb.com/rock.jpg"/>
    <meta property="og:site_name" content="IMDb"/>
    <meta property="fb:admins" content="USER_ID"/>
    <meta property="og:description"
          content="A group of U.S. Marines, under command of
                   a renegade general, take over Alcatraz and
                   threaten San Francisco Bay with biological
                   weapons."/>

    ...
  </head>
  ...
</html>
Yale University
From Wikipedia, the free encyclopedia

"Yale" redirects here. For other uses, see Yale (disambiguation).

Yale University is a private Ivy League university located in New Haven, Connecticut. Founded in 1701 in the Colony of Connecticut, the university is the third-oldest institution of higher education in the United States. Yale has produced many notable alumni, including five U.S. presidents, seventeen U.S. Supreme Court Justices, and several foreign heads of state.

Incorporated as the Collegiate School, the institution traces its roots to 17th-century clergymen who sought to establish a college to train clergy and political leaders for the colony. In 1718, the College was renamed Yale College to honor a gift from Elihu Yale, a governor of the British East India Company. In 1861, the Graduate School of Arts and Sciences became the first U.S. school to award the Ph.D. [8]

Yale College was transformed, beginning in the 1930s, through the establishment of residential colleges: 12 now exist and two more are planned. Almost all tenured professors teach undergraduate courses, more than 2,000 of which are offered annually. [9][10]

The University's assets include a US$16.3 billion endowment, [11] the second largest of any academic institution, as well as the second largest academic library in the world, with some 12.5 million volumes held in more than two dozen libraries. [12]

Yale and Harvard have been rivals in academics, athletics, and other activities for most of their history, competing annually in The Game and the Harvard-Yale Regatta.

http://en.wikipedia.org/wiki/Yale_University
1. Fitzgerald, F. Scott (1920). *This Side of Paradise*, chapter 2: “half-a-dozen seats were kept from sale and occupied by six of the worst-looking vagabonds that could be hired from the streets... At the moment in the show when Firebrand, the Pirate Chief, pointed at his black flag and said, ‘I am a Yale graduate’ – note my skull and bones!” – at this very moment the six vagabonds were instructed to rise conspicuously and leave the theatre with looks of deep melancholy and an injured dignity. It was claimed though never proved that on one occasion the hired Elis were swallowed by one of the real things.


Non-controversial edits

Editors who may have a conflict of interest may nevertheless add material that accords with the Photographs and media files or Subject and culture sector professionals sections below, and are allowed to make certain kinds of non-controversial edits, such as:

1. Removing spam and reverting vandalism.
2. Deleting content that violates Wikipedia's biography of living persons policy.
3. Fixing spelling and grammar errors.
4. Reverting or removing their own COI edits. Cleaning up your own mess is allowed and encouraged.
5. Making edits that have been agreed to on the talk page.
6. Adding citations, especially when another editor has requested them.

If another editor objects for any reason, then it's a controversial edit. Such edits should be discussed on the article's talk page.

Photographs and media files

Wikimedia Commons encourages parties with potential conflicts of interest to upload digital media files, such as photographs, illustrations, audio files, and video clips, so long as the media is of good quality, is in a format we use, and the copyright holder is willing to do so under one of the free licenses we accept.

While Commons prefers full resolution media, reduced resolution images are acceptable when the copyright owner is unwilling to freely license a full quality image. See Commons:Welcome for detailed requirements.

Once media files are uploaded to Commons, they can then be incorporated into Wikipedia articles where appropriate. The best approach is to mention the availability of the image or media files on the article's talk page. But it is usually acceptable to edit the article directly to add one or two images that illustrate the existing article content, e.g., adding a publicity mug shot to the biography of a performing artist (but only if it's available under a free license).

Subject and culture sector professionals

Museum curators, librarians, archivists, art historians, heritage interpreters, conservators, documentation managers, subject specialists, and managers of a special collection (or similar profession) are encouraged to use their knowledge to help improve Wikipedia, or to share their information with Wikipedia in the form of links to their resources.

Wikipedia prefers the addition of content using reliable sources. If a link cannot be used to expand an article, the link may be placed under further reading or external links providing the link complies with the external links guideline and is not covered by any of the points under links to avoid. Keep in mind that Wikipedia is not a mirror or a repository of links, images, or media files.

Don't create a shared organizational account and don't include the name of the organization in the account name. It is recommended but not required for such editors to declare their affiliation on their user page.

When in doubt, first suggest the addition of the link on the talk page of the article, or discuss the addition with an appropriate WikiProject. For more information, see Advice for the cultural sector.
Non-controversial edits

Editors who may have a conflict of interest may nevertheless add material that accords with the Photographs and media files or Subject and culture sector professionals sections below, and are allowed to make certain kinds of non-controversial edits, such as:

1. Removing spam and reverting vandalism.
2. Deleting content that violates Wikipedia's biography of living persons policy.
3. Fixing spelling and grammar errors.
4. Reverting or removing their own COI edits. Cleaning up your own mess is allowed and encouraged.
5. Making edits that have been agreed to on the talk page.
6. Adding citations, especially when another editor has requested them.

If another editor objects for any reason, then it's a controversial edit. Such edits should be discussed on the article's talk page.

Photographs and media files

Wikimedia Commons encourages parties with potential conflicts of interest to upload digital media files, such as photographs, illustrations, audio files, and video clips, so long as the media is of good quality, is in a format we use, and the copyright holder is willing to do so under one of the free licenses we accept.

While Commons prefers full resolution media, reduced resolution images are acceptable when the copyright owner is unwilling to freely license a full quality image. See Commons:Welcome for detailed requirements.

Once media files are uploaded to Commons, they can then be incorporated into Wikipedia articles where appropriate. The best approach is to mention the availability of the image or media files on the article's talk page. But it is usually acceptable to edit the article directly to add one or two images that illustrate the existing article content, e.g., adding a publicity mug shot to the biography of a performing artist (but only if it's available under a free license).

Subject and culture sector professionals

Museum curators, librarians, archivists, art historians, heritage interpreters, conservators, documentation managers, subject specialists, and managers of a special collection (or similar profession) are encouraged to use their knowledge to help improve Wikipedia, or to share their information with Wikipedia in the form of links to their resources.

Wikipedia prefers the addition of content using reliable sources. If a link cannot be used to expand an article, the link may be placed under further reading or external links providing the link complies with the external links guideline and is not covered by any of the points under links to avoid. Keep in mind that Wikipedia is not a mirror or a repository of links, images, or media files.

Don't create a shared organizational account and don't include the name of the organization in the account name. It is recommended but not required for such editors to declare their affiliation on their user page.

When in doubt, first suggest the addition of the link on the talk page of the article, or discuss the addition with an appropriate WikiProject. For more information, see Advice for the cultural sector.
linkypedia helps you **explore** how your web content is used on wikipedia, **extract** crowd-sourced information about your web resources, and **connect** with wikipedians who are using your stuff. More about the rationale for linkypedia can be found [here](http://en.wikipedia.org/wiki/Yale_University). Note: this is a demo site, and can go down at any time. But please feel free to [download](http://en.wikipedia.org/wiki/Yale_University) the code and run it in your own environment if you want.

**recently found:** Talk:Spencer Brown (rugby player born 1921) referencing NLA Australian Newspapers

- Internet Archive: Free Movies, Music, Books & Wayback Machine (33,829)
  - bioguide.congress.gov (19,984)
  - FIFA.com (18,703)
- WorldCat.org: The World’s Largest Library Catalog (17,352)
  - SAO/NASA ADS: ADS Home Page (17,052)
  - Free eBooks by Project Gutenberg (14,344)
  - arXiv.org e-Print archive (9,631)
  - Wikispecies, free species directory (9,621)
- Welcome to the United Nations: It’s Your World (9,508)
  - American Memory from the Library of Congress (8,334)
  - Library of Congress (4,871)
  - Library of Congress Handle Proxy (4,579)
  - World Wide Web Consortium (W3C) (4,427)
  - THOMAS (Library of Congress) (3,641)
- National Archives and Records Administration (2,765)
  - Te Ara Encyclopedia of New Zealand (2,199)
linkypedia helps you **explore** how your web content is used on wikipedia, **extract** crowd-sourced information about your web resources, and **connect** with wikipedians who are using your stuff. More about the rationale for linkypedia can be found [here](http://en.wikipedia.org/wiki/Yale_University). Note: this is a demo site, and can go down at any time. But please feel free to [download](http://en.wikipedia.org/wiki/Yale_University) the code and run it in your own environment if you want.

**recently found:** Talk:Spencer Brown (rugby player born 1923) referencing [NLA Australian Newspapers](http://www.yale.edu/)

- [Internet Archive: Free Movies, Music, Books & Wayback Machine](http://www.yale.edu/) (33,829)
  - [bioguide.congress.gov](http://www.yale.edu/) (19,984)
  - [FIFA.com](http://www.yale.edu/) (18,703)
- [WorldCat.org: The World’s Largest Library Catalog](http://www.yale.edu/) (17,352)
  - [SAO/NASA ADS: ADS Home Page](http://www.yale.edu/) (15,052)
  - [Free eBooks by Project Gutenberg](http://www.yale.edu/) (14,344)
  - [arXiv.org e-Print archive](http://www.yale.edu/) (9,631)
  - [WikiSpecies, free species directory](http://www.yale.edu/) (9,621)
- [Welcome to the United Nations: It’s Your World](http://www.yale.edu/) (9,508)
  - [American Memory from the Library of Congress](http://www.yale.edu/) (8,334)
  - [Library of Congress](http://www.yale.edu/) (4,871)
  - [Library of Congress Handle Proxy](http://www.yale.edu/) (4,579)
  - [World Wide Web Consortium (W3C)](http://www.yale.edu/) (4,427)
  - [THOMAS (Library of Congress)](http://www.yale.edu/) (3,641)
  - [National Archives and Records Administration](http://www.yale.edu/) (2,767)
  - [Te Ara Encyclopedia of New Zealand](http://www.yale.edu/) (2,192)
# Summary Information for Yale University

<table>
<thead>
<tr>
<th>Website Title</th>
<th>Yale University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Website URL</td>
<td><a href="http://www.yale.edu/">http://www.yale.edu/</a></td>
</tr>
<tr>
<td>Wikipedia Links</td>
<td>703</td>
</tr>
<tr>
<td>Wikipedia Pages</td>
<td>556</td>
</tr>
<tr>
<td>Wikipedia Categories</td>
<td>0</td>
</tr>
<tr>
<td>Last Crawl</td>
<td></td>
</tr>
<tr>
<td>Added By</td>
<td>127.0.0.1</td>
</tr>
</tbody>
</table>

*If you like the site, want to add your own features, or want to run an instance yourself, please checkout the code from GitHub.*
## Wikipedia Pages for Yale University

Wikipedia pages for Yale University. Put this feed in your favorite feed reader and get notified of new pages that reference this website.

<table>
<thead>
<tr>
<th>Wikipedia Page</th>
<th>Links</th>
<th>Last Update at Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Yale University student organizations</td>
<td>13</td>
<td>2010-08-11 01:08:38</td>
</tr>
<tr>
<td>Yale University</td>
<td>28</td>
<td>2010-08-12 09:08:52</td>
</tr>
<tr>
<td>Residential colleges of Yale University</td>
<td>24</td>
<td>2010-07-27 05:07:44</td>
</tr>
<tr>
<td>Soviet–German relations before 1941</td>
<td>20</td>
<td>2010-08-12 10:08:23</td>
</tr>
<tr>
<td>Talk:Tripartite Pact/Archive 2</td>
<td>20</td>
<td>2010-08-07 07:08:34</td>
</tr>
<tr>
<td>Wikipedia:Dead external links/301y</td>
<td>16</td>
<td>2010-03-06 10:03:56</td>
</tr>
<tr>
<td>Territorial evolution of the United States</td>
<td>13</td>
<td>2010-08-12 02:01:08</td>
</tr>
<tr>
<td>Non-native territorial evolution of North America 1763 to present</td>
<td>10</td>
<td>2010-07-27 09:07:26</td>
</tr>
<tr>
<td>User:JFriedmann</td>
<td>10</td>
<td>2010-07-15 12:07:01</td>
</tr>
<tr>
<td>Soviet occupation of Bessarabia and Northern Bukovina</td>
<td>8</td>
<td>2010-08-10 01:08:30</td>
</tr>
<tr>
<td>Talk:History of Poland (1930–1945)</td>
<td>8</td>
<td>2010-07-31 06:07:27</td>
</tr>
<tr>
<td>Talk:Treaty of Tripoli/Archive 1</td>
<td>7</td>
<td>2010-07-27 09:07:42</td>
</tr>
<tr>
<td>Yale Political Union</td>
<td>6</td>
<td>2010-07-17 08:07:29</td>
</tr>
<tr>
<td>Environmental Sustainability Index</td>
<td>6</td>
<td>2010-07-26 04:07:07</td>
</tr>
<tr>
<td>Talk:Yale University/Archive 4</td>
<td>6</td>
<td>2010-01-26 03:01:07</td>
</tr>
</tbody>
</table>
Wikipedia categories for Yale University.

<table>
<thead>
<tr>
<th>Wikipedia Categories</th>
<th>Wikipedia Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>All articles with unsourced statements</td>
<td>382</td>
</tr>
<tr>
<td>Living people</td>
<td>126</td>
</tr>
<tr>
<td>All articles needing additional references</td>
<td>108</td>
</tr>
<tr>
<td>All articles with dead external links</td>
<td>97</td>
</tr>
<tr>
<td>All pages needing cleanup</td>
<td>96</td>
</tr>
<tr>
<td>All articles to be expanded</td>
<td>67</td>
</tr>
<tr>
<td>All articles with specifically-marked weasel-worded phrases</td>
<td>55</td>
</tr>
<tr>
<td>Yale University faculty</td>
<td>50</td>
</tr>
<tr>
<td>Articles containing Russian language text</td>
<td>46</td>
</tr>
<tr>
<td>Yale University</td>
<td>46</td>
</tr>
<tr>
<td>Articles with unsourced statements from May 2000</td>
<td>45</td>
</tr>
<tr>
<td>Articles with unsourced statements from February 2010</td>
<td>38</td>
</tr>
<tr>
<td>Articles with unsourced statements from July 2008</td>
<td>37</td>
</tr>
<tr>
<td>Articles with unsourced statements from February 2007</td>
<td>36</td>
</tr>
<tr>
<td>Articles with unsourced statements from April 2010</td>
<td>35</td>
</tr>
<tr>
<td>All articles lacking in-text citations</td>
<td>34</td>
</tr>
<tr>
<td>Articles with unsourced statements from May 2008</td>
<td>34</td>
</tr>
</tbody>
</table>
The following Wikipedia users have mentioned Yale University in their Wikipedia profile, or associated user pages. The plan is to eventually also include users who have contributed specific links to Wikipedia articles.

<table>
<thead>
<tr>
<th>Username</th>
<th>Registered</th>
<th>Edit Count</th>
<th>Emailable</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>3ala2addy</td>
<td>Aug. 4, 2009, 11:31 a.m.</td>
<td>32</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>68.230.82.231</td>
<td>None</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGK</td>
<td>Feb. 27, 2006, 4:51 p.m.</td>
<td>26673</td>
<td>✓</td>
<td>sysop</td>
</tr>
<tr>
<td>Adam63</td>
<td>Jan. 20, 2007, 3:44 p.m.</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ahmadi</td>
<td>Jan. 19, 2007, 5:56 p.m.</td>
<td>1802</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AlexR</td>
<td>Dec. 26, 2002, 4:34 p.m.</td>
<td>3829</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Amorken</td>
<td>Jan. 15, 2007, 9:23 p.m.</td>
<td>6</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Anggarrgoon</td>
<td>Jan. 13, 2006, 2:33 a.m.</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antony-22</td>
<td>Nov. 8, 2006, 3:02 p.m.</td>
<td>2342</td>
<td>✓</td>
<td>reviewer</td>
</tr>
<tr>
<td>ArmyAnt</td>
<td>Feb. 5, 2007, 8:32 p.m.</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arsonal</td>
<td>Jan. 18, 2006, 3:37 a.m.</td>
<td>8153</td>
<td>✓</td>
<td>autoreviewer reviewer rollbacker</td>
</tr>
<tr>
<td>Baler</td>
<td>April 25, 2004, 2:16 a.m.</td>
<td>12675</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>BS2011</td>
<td>May 30, 2008, 7:50 p.m.</td>
<td>105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bengalski</td>
<td>July 8, 2005, 2:44 p.m.</td>
<td>875</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bkonrad</td>
<td>Feb. 13, 2004, 3:50 p.m.</td>
<td>80126</td>
<td>✓</td>
<td>sysop</td>
</tr>
<tr>
<td>Blue-Haired Lawyer</td>
<td>March 18, 2006, 10:27 p.m.</td>
<td>5136</td>
<td>✓</td>
<td>reviewer</td>
</tr>
<tr>
<td>Blur4760</td>
<td>Feb. 8, 2006, 7:44 p.m.</td>
<td>191</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caribbean H.O.</td>
<td>Dec. 12, 2006, 2:52 p.m.</td>
<td>24100</td>
<td>✓</td>
<td>sysop</td>
</tr>
</tbody>
</table>
Conclusion

http://www.flickr.com/photos/photograham/155636430/
Thank You

mark@matienzo.org
http://matienzo.org/
twitter: @anarchivist