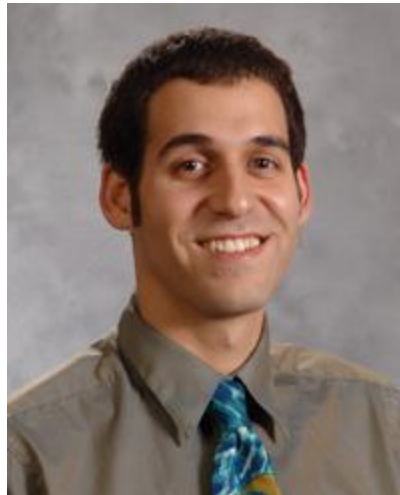


OCLC Research Webinar

# Wikipedia and Libraries: What's the Connection?



Featuring Max Klein,  
Wikipedian in Residence  
31 July 2012

#orwikipedia  
@notconfusing

**IMPORTANCE**

# SIZE OF WIKIPEDIA

## Top Largest Websites

1. Google
2. Facebook
3. Youtube
4. Yahoo!
5. Baidu
6. Wikipedia
7. Windows Live!
8. QQ
9. Amazon
10. LinkedIn

>500,000,000  
Monthly Views

6<sup>th</sup> Largest Site

89,000 Active  
Editors

# OF WHICH ARE NONPROFIT

## Top Largest Websites

1. Google
2. Facebook
3. Youtube
4. Yahoo!
5. Baidu
6. Wikipedia
7. Windows Live!
8. QQ
9. Amazon
10. LinkedIn

Only  
Top 100 site  
without  
user  
surveillance

**CONNECTION**

# MISSIONS

“Imagine a world in which every single human being can freely share in the sum of all knowledge. That's our commitment.”

Library or Wikipedia?

Complete

Free

Access

# QUALITY

Wikipedia had a reputation for being unreliable but Wikipedians are dedicated to quality sourcing

The latest data shows users rating scores in this order

1. “Well-sourced”
2. “Readable”
3. “Complete”
4. “Neutral”

And all are rising over time

[1]

[http://en.wikipedia.org/wiki/User:Protonk/Article\\_Feedback](http://en.wikipedia.org/wiki/User:Protonk/Article_Feedback)

Highest article rating is “well sourced”[1]

Improving over time

# CITATIONS

- 411,274 citations of books
- 244,236 citations of journals
- 57,868 citations of encyclopedias
- 342,470 of newspapers
- **1,055,845** total **print citations**<sup>[1]</sup>
- **1,169,495** citations of **web**<sup>[1]</sup>

[1]

<http://en.wikipedia.org/wiki/User:Maximilianklein/Citations>

[2]

[http://en.wikipedia.org/wiki/Category:Articles\\_lacking\\_sources](http://en.wikipedia.org/wiki/Category:Articles_lacking_sources)

Print citations as  
common as  
web<sup>[1]</sup>

94.2% contain at  
least one  
reference<sup>[2]</sup>



# CITATIONS

Even in Academia researchers are increasingly using Wikipedia as a source and pedagogical tool.

Wikipedia has been cited 3,679 times in the WoS and Scopus databases during the last nine years[1].

[1]

<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3492/3031>

2003: *Scopus*  
articles cited  
Wikipedia 4  
times

2010: *Scopus*  
articles cited  
Wikipedia 933  
times

# ENGAGEMENT

# RESEARCH FLOW, PRE-INTERNET



# RESEARCH FLOW, NOW



# RESEARCH FLOW, FUTURE



# RESEARCH FLOW, FUTURE



# LINKING WIKIPEDIA TO LIBRARIES

It's not contentious that research should end at the sources.

The open question is how to create the links between Wikipedia and Libraries.

Who's linking?

What needs linking?

Where to link?

# COLLABORATION SO FAR

- **Institutions** have edited Wikipedia on their own special collections.
- Editathons bring **Wikipedians** to the sources.
- Workshops train **new editors** on how to create edit Wikipedia.
- Bots (robots) are **software** that automatically edit Wikipedia.
- Media Donations are **bulk uploads** of media to *the commons*.

Direct Editing

Editathons

Workshops

Bots

Media Donation



# DIRECT EDITING

The model of Librarians editing Wikipedia and linking to their own special collections was probably the earliest attempt at linking.

**For instance:** University of Washington Libraries [1].

**Advantages:** can be completed by a paid workforce.

**Disadvantages:** Wikipedia community can be hostile to *conflict of interest editors* if they (unknowingly) break Wikipedia policy.

[1]

<http://interactivearchivist.archivists.org/case-studies/wikipedia-at-uw/#lessons>

Link traffic to special collections sites

Done by paid staff

Wikipedia community can be hostile if not executed perfectly

# EDITATHONS

Invite Wikipedians and patrons alike into the library for an event. Possibly giving access to rare materials.

**For instance:** Princeton Libraries[1].

**Advantages:** Brings newcomers to the library. Exposes rare materials.

**Disadvantages:** Varying quantity and quality of work. Requires event planning.

[1]

[http://outreach.wikimedia.org/w/index.php?title=File:Q\\_VLL\\_talk\\_slideshow.pdf](http://outreach.wikimedia.org/w/index.php?title=File:Q_VLL_talk_slideshow.pdf)

Planned events

Brings foot traffic

Expose rare materials

Variable results

# WORKSHOPS

Encouraging newcomers to edit by hosting workshops on editing techniques.

**For instance:** Harvard Workshop<sup>[1]</sup>.

**Advantages:** Focus on new editors can bring in otherwise hesitant participants.

**Disadvantages:** Does not ensure ensuing work.

[1]

[http://en.wikipedia.org/wiki/Wikipedia:WLL\\_at\\_Harvard](http://en.wikipedia.org/wiki/Wikipedia:WLL_at_Harvard)

Garner new participants

Low inertia

No guaranteed action

# BOTS

(Ro)bots are software that can edit Wikipedia extensively.

**For instance:** VIAFbot[1].

**Advantages:** Large impact if the problem can be solved algorithmically.

**Disadvantages:** Requires very specific data focus and skills.

[1]

<http://en.wikipedia.org/wiki/Wikipedia:VIAF>

Scalable

Useful for data integration projects

Large community approval process

Requires programming skills

# MEDIA DONATION

Donations of Public Domain (or CC-BY-SA) images, and texts  
*Wikimedia commons* help support free online materials.

**For instance:** National Archives[1].

**Advantages:** Creates promotional press. Can enrich media by unforeseeable reuse.

**Disadvantages:** Does not link back to donating institution.

[1]

<http://en.wikipedia.org/wiki/Wikipedia:NARA>

Generates good press

Is often remixed

Does not necessarily create links

# LIBRARIAN OPPORTUNITIES

**Reliable sources noticeboard** is a discussion forum on assessing the quality of sources[1].

**“Category:Unreferenced”** is a list of articles lacking sourcing entirely[2].

**The Teahouse** is a new, friendly approach to peer support on Wikipedia[3].

[1]

[http://en.wikipedia.org/wiki/Wikipedia:Reliable\\_sources/Noticeboard](http://en.wikipedia.org/wiki/Wikipedia:Reliable_sources/Noticeboard)

[2]

[http://en.wikipedia.org/wiki/Category:Articles\\_lacking\\_sources](http://en.wikipedia.org/wiki/Category:Articles_lacking_sources)

[3]

<http://en.wikipedia.org/wiki/Wikipedia:Teahouse>

Give advice on references

Improve articles

Grow an inside perspective

**FORENSICS**

# THE TALK PAGE

Sometimes known as the “discussion page” this is the location for editors to talk about the article.

Duality: every page is two pages.

Meta-page

Permanently  
attached

Talk to other  
editors



# THE TALK PAGE – EXAMPLES

Egyptian Revolution

What would be some of the major talk today around the Egyptian Revolution article?

[http://en.wikipedia.org/wiki/2011\\_Egyptian\\_revolution](http://en.wikipedia.org/wiki/2011_Egyptian_revolution)

# PAGE RATINGS

## Quality

Featured Article

A

Good Article

B

C

Start

Stub

## Importance

Top

High

Mid

Low

Bottom

No

# PAGE RATINGS – EXAMPLES

Egyptian Revolution

What score would you guess?

[http://en.wikipedia.org/wiki/Talk:2011\\_Egyptian\\_revolution](http://en.wikipedia.org/wiki/Talk:2011_Egyptian_revolution)

[1\\_Egyptian\\_revolution](http://en.wikipedia.org/wiki/Talk:2011_Egyptian_revolution)

What percentage are stubs?

[http://en.wikipedia.org/wiki/Wikipedia:Version\\_1.0\\_Editorial\\_Team/Assessment#Quality\\_scale](http://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Assessment#Quality_scale)

# VIEW HISTORY

The entire revision history for a page is always stored on Wikipedia.

That also includes talk pages.

Complete History

For Every Page

Forever

# VIEW HISTORY- EXAMPLES

Egyptian Revolution 2011

How soon after the first protest  
did the first draft come?

[http://en.wikipedia.org/w/index.php?title=Talk:2011\\_Egyptian\\_revolution&action=history](http://en.wikipedia.org/w/index.php?title=Talk:2011_Egyptian_revolution&action=history)

# USER PAGES

The user page itself is the public face that the user chooses to show.

The attached talk page of a user page acts as the message system.

Private messaging is impossible by design.

Profile

Messaging System

Transparency

# USER PAGES—EXAMPLES

User:The Egyptian Liberal

What was s/he discussing?

[http://en.wikipedia.org/wiki/User:The\\_Egyptian\\_Liberal](http://en.wikipedia.org/wiki/User:The_Egyptian_Liberal)

# WHAT LINKS HERE

Part of Wikipedia's special structure is that links are **bidirectional**.

We take links forwards all the time.  
Why not backwards?

All pages

Counting tool

New browsing  
mode



# WHAT LINKS HERE– EXAMPLES

## Pages

What do you think links to Egyptian Revolution?

[http://en.wikipedia.org/wiki/Special:WhatLinksHere/2011\\_Egyptian\\_revolution](http://en.wikipedia.org/wiki/Special:WhatLinksHere/2011_Egyptian_revolution)

## Templates

Infoboxes are a type of template.

They are the boxes at the right with structured data

What about “infobox book”

[http://en.wikipedia.org/wiki/Template:Infobox\\_book](http://en.wikipedia.org/wiki/Template:Infobox_book)

# PAGE VIEWS

Simple page view counter displayed over time.

All pages

All languages

Up to 90 days  
prior

# PAGE VIEWS – EXAMPLES

## Pages

On the July 18<sup>th</sup> release of a new Joyce Biography, what effect on Ulysses?

[http://stats.grok.se/en/latest/Ulysses\\_\(novel\)](http://stats.grok.se/en/latest/Ulysses_(novel))

**THEORY**

# FIVE PILLARS

An  
Encyclopedia

Neutral  
Point of View

Free  
Content

Civility

No Firm  
Rules

[http://en.wikipedia.org/wiki/Wikipedia:Five\\_pillars](http://en.wikipedia.org/wiki/Wikipedia:Five_pillars)

# ZEROETH LAW OF WIKIPEDIA

“The problem with Wikipedia is that it only works in practice. In theory, it can never work.” ~Anon

**VIAFBOT**

# VIAFBOT - IDEA

VIAF – the **V**irtual **I**nternational **A**uthority **F**ile – is a matching algorithm for over 20 national name authority files.

In VIAF's algorithmic matching process it uses Wikipedia as to help disambiguate creators.

When a successful match is found and Wikipedia is used, VIAF links to the Wikipedia article it used.

This VIAF->Wikipedia link could be reciprocated on Wikipedia, making it part of a web of services.

An aggregating authority file

Uses and links to Wikipedia data

Would be valuable if Wikipedia reciprocated links



# VIAFBOT - PROCESS

VIAFbot proposes to edit 260,000 articles.

Getting approval on large projects on Wikipedia is difficult because there is no central office.

Instead you must find *community consensus*.

Phases:

- Village Pump (discussion)
- Request for Comment (officials)
- Bot Approval Group (review)

We're close to doing it [1].

[1]

[http://en.wikipedia.org/wiki/Wikipedia:Authority\\_control\\_integration\\_proposal](http://en.wikipedia.org/wiki/Wikipedia:Authority_control_integration_proposal)

260,000 articles to edit

Requires Approval

No office to call

Must find community consensus

>30,000 community members

# VIAFBOT - EXAMPLES

VIAF entry for Maya Angelou links to Wikipedia.

[http://viaf.org/viaf/7386077/#Angelou,  
\\_Maya](http://viaf.org/viaf/7386077/#Angelou,_Maya)

VIAF entries for Irvine Welsh

[http://viaf.org/viaf/101956084/#Wels  
h,\\_Irvine](http://viaf.org/viaf/101956084/#Welsh,_Irvine)

Once linked we could use the linked data to automatically glean more facts.

<http://viaf.org/viaf/7386077/rdf.xml>

**IDEAS**

# REQUEST FOR COMMENT

“*Wikipedia Loves Libraries*” is a  
Wikipedian-centered approach  
to the two collaborating.

We want more Library voices and  
ideas.

What can you imagine in this space?

Formally <2  
years old

More library  
voices can shape  
projects

# COMMENTS AND QUESTIONS

More online at:

[http://outreach.wikimedia.org/wiki/Wikipedia\\_Loves\\_Libraries](http://outreach.wikimedia.org/wiki/Wikipedia_Loves_Libraries)

OCLC Research Webinar

## Wikipedia and Libraries: What's the Connection?

**#orwikipedia**

**Max Klein**

kleinm@oclc.org

@notconfusing

**OCLC Research**

<http://www.oclc.org/research/>

**Our Webinars**

<http://www.oclc.org/research/events/webinars.htm>



**OCLC Research**

Exploration, innovation and community  
for libraries and archives.