Contribution Patterns Among Active Wikipedians:
Finding and Keeping Content Creators

Seth Anthony ( [[User:Seth Ilyys]] )
Wikimania 2006 – Wiki Community Dynamics
5 August 2006
Sociology of Online Communities

Social sciences suffer from:
- paucity of data
- non-quantitative data
- skepticism

(Not that non-quantitative research is invalid.)

Online communities provide a rich data resource, Wikipedia in particular.
A Confession and a Shameless Plug

Hi, I'm Seth, and I'm a wikiholic!

My wikihistory....
- began editing en.wikipedia in December 2003
- en.wikipedia admin since March 2004
- too many article namespace edits

But what have I done?
- Only one FA: [[Caroline Island]]
  (just promoted this week)
Central Questions

As of this morning, en.wikipedia has:
  almost 1,300,000 articles
  500,000,000 words
Who wrote all this?
  1,900,000 registered accounts
  38,000 editors (>5 edits, Jun '06)
  70,000,000 edits
How do we identify them?
How do we value them? (Are barnstars enough?)
Classifying Individual Edits

Edits with negative value:
- Vandalism/”trolling”

Edits without content value (but still necessary!):
- Administration/bureaucracy
- Discussion/dispute resolution

Edits with content value:
- Stylistic modification of articles
- Addition of content to articles – our interest
Muhammad ibn Abd-al-Wahhab

From Wikipedia, the free encyclopedia

Revision as of 17:37, 2 August 2006 (edit)
Zaydi-Ismoudeh (Talk | contribs | block)
(← Older edit)

Current revision (08:15, 5 August 2006) (edit)
Striver (Talk | contribs | block) [rollback]

Muhammad ibn Abd-al-Wahhab

al-Tamimi (1703AD – 1792AD) (Arabic: محمد بن عبد الوهاب التميمي) was an Arab theologian born in the Najd, in present-day Saudi Arabia and the most famous scholar of the movement within Islam known as the Salafi movement.

Contents
1 Legacy
2 Biography
2.1 Childhood
CSS Georgia (battery)

CSS Georgia, also known as State of Georgia and Ladies' Ram, was an ironclad floating battery built at Savannah, Georgia in 1862–1863. Placed under command of Lieutenant Washington Gwathmey, CSSN, she was employed in defending the river channels below Savannah, training her batteries against the Union advance. Since she lacked effective locomotive power, the Confederates found it necessary to fire and destroy her during the evacuation of Savannah on 21 December 1864.

This article includes text from the public domain Dictionary of American Naval Fighting Ships.
Who are our content creators?

We don't think about this much, but we bemoan the loss of the stressed admin. Is this valid/productive?

Hypothesis -- Content creators aren't (generally):
- admins - too busy administrating
- the “long tail” - edits typically minor

They are:
- “pre-admins” - before caught in wikiocracy
- non-ambitious, occasional editors (specialists?)
“Thousands of editors at thousands of keyboards...”

Sample of 250 recent edits:
- Outside article namespace: 28%
- Article talk namespace: 10%
  (potentially content-productive)
- Vandalism: 5%
- “Tweaking”: 45%
  (small changes/stylistic)
- Substantive changes/additions: 10%
- Creation of new articles: 2%
Who creates content?

Of high content edits: (32 in this set)
- 0 were made by an admins
- 69% are made by someone with a username
- 52% by someone with a userpage
- none have a barnstar

Median date of first edit: April 2006
  (only 3-4 months ago)
Median number of edits: 200 (range: 1-5000)
- Span of 50 edits:
  First 50 took 400 hours (77% in article nsp.)
  Recent 50 took 90 hours (76% in article nsp.)
What are our admins doing?

Random sample of 28 en admins (Jul '06):
Most recent 50 edits:
- How often?
  50 edits took them from 8 to 1300 hours
  median: most recent 50 edits took 150 hours
  (50 edits/6 days or 1 edit/3 hours)
- Consistency?
  recent 500 edits took them 1219 hours
  (10x edits in 8x time)
- Where?
  27 out of 50 edits (57%) in article namespace
They weren't always like this...

Early in their Wikipedia careers, they:

Edited less frequently:
  First 50 edits took a median 440 hours (18 days)
  (compare: median 150 hours for recent edits)

Created content more:
  38 of first 50 edits (76%) in article namespace
  (compare 57% for recent edits)

Edited less consistently:
  First 500 edits took 1650 hours
  (compare above: 10x edits in 4x time)
Implications for Wikipedia...

Content creators seem to be:
- occasional and less frequent editors
- may be specialists (subject matter knowledge)

Admins:
- former content creators, now janitors

Wikipedia goals:
- quality vs. quantity – different contributor sets?
- how do we value content contributors?
Future directions and acknowledgements

Assistance wanted with data mining...
We need robust statistics on editing patterns:
  - across projects, languages/cultures, time
Comparisons with other volunteer-driven efforts:
  - non-profit organizations, political campaigns

Thanks to:
- Wikipedia contributors, readers, and critics
- Interiot and other tool creators
- Colorado State University
Licensing

This file is licensed under the terms of the

GNU Free Documentation License:
http://www.gnu.org/copyleft/fdl.html

and

the Creative Commons-BY-SA license:
http://creativecommons.org/licenses/by-sa/2.5/