Flagged revisions

Latest results
Summary

- Goals
- Methodology
- Results
- Questions and feedback
Goals

- Study the **impact of flagged revisions** in the editorial work of the **German Wikipedia**.
  - Focus on **anonymous** editors.
  - Focus on **vandalism** and reverts.

- Questions:
  - Is the vandalism from anonymous reduced?
  - Is the number of anonymous blocks reduced?
  - Does it discourage anonymous from editing?
  - Does it decrease number of requests for new accounts? (TODO)
Methodology (overview)

- Parse `page-logging.xml`
- EDA on the data set.
  - Survival analysis on time to review/revert an edit.
  - Numbers and time series:
    - Reverts
    - Blocks /IP-blocks
    - (Semi)-protection
    - **Sighted** anonymous edits
- Comparison with similar versions.
  - **Polish & Russian** (also with flagged-revs).
  - Also WKP comparable size (FR, IT)
Methodology (concepts)

- We focus on anonymous contributions.

- Sighting actions (manual)
  - **Approved**: Revisions *manually* flagged as OK.
  - **Approved-i**: Introduced later to identify first approvals.

- Sighting actions (automated)
  - Approved-a, approved-ia.
  - Automated approval for trusted editors.
  - Filtered-out.
At first, all revisions in a given series (same page) are pending.
Methodology (approvals)

Then, we mark all revisions *directly approved*.

SET OF CONSECUTIVE ANON REVS (page X)

- $r_0$: Pending
- $r_1$: Pending
- $r_2$: Approved-direct
- $r_3$: Pending
- $\ldots$
- $r_{N-1}$: Approved-direct
- $r_N$: Pending
Methodology (approvals)

SET OF CONSECUTIVE ANON REVS (page X)

\[ r_0 \] Pending

\[ r_1 \] Pending

\[ r_2 \] Approved-direct

\[ r_3 \] Approved-implicit

\[ \ldots \]

\[ r_{N-1} \] Approved-direct

\[ r_N \] Pending

All revs in between 2 app-direct are also approved implicitly
Methodology (approvals)

SET OF CONSECUTIVE ANON REVS (page X)

\[ r_0 \] Approved-implicit

\[ r_1 \] Approved-implicit

\[ r_2 \] Approved-direct

\[ r_3 \] Approved-implicit

\[ \ldots \]

\[ r_{N-1} \] Approved-direct

\[ r_N \] Pending

All revs before direct approval are also approved implicitly.
Methodology (reverts)

- **Vandal** reverts
  - Identified by regexps.
  - Does not include standard admin reverts.
- **Other** reverts.
  - Reverts without explicit reference to vandalism
  - Includes admin reverts.
Methodology (reverts)

- Detecting reverted revisions
  - First, look for the newest revision with the same size (after revert action).
  - From that on, mark as reverted all consecutive revisions performed by same IP.
  - If that fails, look for IP info in the comment field, and look for newest revision performed by that IP.
  - Then, same procedure to mark all consecutive revisions.
- Feedback?
RESULTS
Evolution anonymous edits

Number of contributions from anonymous users

Time

2004 2005 2006 2007 2008 2009 2010

Low

High

High
Evol anonymous edits (focus)

Number of contributions from anonymous users

Time

Num. of contributions

- 2007.0
- 2007.5
- 2008.0
- 2008.5
- 2009.0
- 2009.5
- 2010.0

DE
Evolution review actions
Time to approve/revert revisions

SET OF CONSECUTIVE ANON REVS (page X)

- $r_0$: Logged user
- $r_1$: Anon-user
- $r_2$: Logged user
- $r_3$: Logged user
- $r_{N-1}$: Logged user → Approve $r_1$
- $r_N$: Logged user

$tstamp(r_{N-1}) - tstamp(r_1)$
Time to approve/revert revisions

Revisions approved or Reverted at a very fast pace
“Truth in numbers”

- Revert actions performed at very fast pace.
  - **Revert** (median): 48 min.
  - **Revert-v** (median): **36 seg**.

- High number of actions registered → **accuracy**

<table>
<thead>
<tr>
<th>Status</th>
<th>events</th>
<th>*rmean</th>
<th>*se(rmean)</th>
<th>median</th>
<th>0.95LCL</th>
<th>0.95UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2892231</td>
<td>241.8</td>
<td>0.578</td>
<td>3.369</td>
<td>3.345</td>
<td>3.3933</td>
</tr>
<tr>
<td>2</td>
<td>1839583</td>
<td>988.0</td>
<td>1.442</td>
<td>194.082</td>
<td>193.445</td>
<td>194.7781</td>
</tr>
<tr>
<td>3</td>
<td>31774</td>
<td>51.0</td>
<td>0.744</td>
<td>0.807</td>
<td>0.756</td>
<td>0.8700</td>
</tr>
<tr>
<td>4</td>
<td>119632</td>
<td>10.9</td>
<td>0.168</td>
<td>0.010</td>
<td>0.010</td>
<td>0.0103</td>
</tr>
</tbody>
</table>

1 = Approved-d ; 2 = Approved-im ; 3 = revert ; 4 = revert-v
Comments on time to app/revert

- Implications
  - Looks like extremely fast pace for acting on revisions.
  - Community takes this new role very seriously.
  - Provides stronger incentive to watch content even closely.
Evolution % reverted edits

Some reverts in the red line could belong to the black one
Evolution % editors who revert

Development of % reverted pages over monthly editors

- reverts-anon-vandals
- reverts-anon
Evolution blocked users

Development of num. blocked users

- block-user
- block-IP

Time

2005 2006 2007 2008 2009 2010
Evolution of protection actions
Conclusions

- In general, flagged revisions did not affect the anonymous editing.
  - Most revisions got approved very rapidly
- More activity on vandalism reverts.
  - Even faster than approval actions.
- Reduced impact of vandalism.
  - Growing number of reverts.
  - On an increasing number of pages.
- Mandatory comments had much more direct influence.
Q: What did happen at the beginning of 2008 for such a high number of user pages protected?

- A: mass-blocking of open proxies.
  - Creating the user page of blocked IPs with a template and protecting them.

We need patterns for detecting reverts:

- Russian Wikipedia
- Polish Wikipedia

Comments and feedback are very welcome.