

WikiRate

rating Wikimedia

Michael Maggs

Chair, Wikimedia UK

Objective

- Increasing need for chapters to measure impact via SMART metrics/targets
- ‘SMART’ = Specific, Measurable, Attainable, Relevant and Timely
- Want to apply best practice to measure improvements in article and media file quality
- Improvements in *quantity* can easily be measured; improvements in *quality* are much harder.

Current situation

- Manual evaluation of quality by experts or the community
- 'gold standard' of quality evaluation
- but significant drawbacks: expense and lack of scalability

Other work

- *WikiMetrics* WMF
- *WikiTrust* University of California
- *wm_metrics* Tools created and used by WMFR
- *SuggestBot* on en Wiki. User:Nettrom
- *ClueBot NG* on en Wiki. Users Cobi & Crispy1989

Build on this

Proposal

- Tool to measure article and media file quality in an automated way
- Not as good as experts on a single article, but very useful when applied to a larger sample.
- By measuring how outputs vary over time we can see the impact that a chapter has on overall article and media file quality.
- Makes Wikimedia quality amenable to SMART analysis for the first time.

Who would use the tool?

- Chapters and other Wikimedia entities who need to ways to measure their own impacts
- WikiProjects and online groups
- Educationalists
- Individual editors
- Academic and Wiki researchers
- The Foundation's analytics teams

How to create the model

- Automatically measuring quality without context-specific human knowledge is a technically hard problem: needs researchers in machine learning techniques
 - 1. Score some articles manually
 - 2. Measure some article attributes (inputs)
 - 3. Adjust some mathematical functions so that the measured inputs produce automatic outputs which replicate the manual quality scores.

Possible inputs (articles)

- Community-added tags:
 - WikiProject quality assessment scores
 - Categorisations, such as stub, start article etc
 - Awards (Featured article, Good Article)
 - Manually-added problem templates
- Stability over time
- Extent of edit warring
- Length of text
- Number of references
- Links in
- Links out
- Obvious spelling and grammatical errors

Possible inputs (media files)

- Awards (Featured Image/Picture status on Commons and on Wikipedias; Quality Image; Valued Image)
- Metadata
 - Camera type
 - Pre-processing before upload (Photoshop etc)
- Usage on main spaces of other Wikimedia projects
- Sharpness measured over the image
- Image size (pixel count)
- Image size on disk
- Categorisations (comprehensiveness, specificity)

Finally, run the finished model

- Then, run the model on some other (unrated) articles and media files.
- The model outputs its best estimate of their quality

Tell us what you would like

Via the WMUK Wiki, at:

[https://wikimedia.org.uk/wiki/
Technology_Committee/Project_requests/
WikiRate_-_rating_Wikimedia](https://wikimedia.org.uk/wiki/Technology_Committee/Project_requests/WikiRate_-_rating_Wikimedia)

Or email me directly:

michael.maggs@wikimedia.org.uk

Big picture

1. Send information to the FDC
2. Make FDC happy
3. Get grant

Thank you

Michael Maggs

michael.maggs@wikimedia.org.uk