What aren't we measuring?: Methods for quantifying wiki-work

Aaron Halfaker
Grouplens Research
University of Minnesota
halfaker@cs.umn.edu

Dario Taraborelli
Wikimedia Foundation
dtaraborelli@wikimedia.org

Brian Keegan
Center for Technology & Social Behavior
Northwestern University
bkeegan@northwestern.edu

Maryana Pinchuk
Wikimedia Foundation
mpinchuk@wikimedia.org

Andrea Forte
College of Information Science and Technology
Drexel University
aforte@drexel.edu

Mikhil Masli
Grouplens Research
University of Minnesota
masli@cs.umn.edu

R. Stuart Geiger
School of Information
University of California, Berkley
rsg@berkley.edu

WikiSym 12, Aug 2729, 2012, Linz, Austria.
ACM 978-1-4503-1605-7/12/08.

Abstract
Wikis and other open collaboration systems rely on the work of contributors to survive. But what is work and how do we quantify it? Answering this question in the right context is essential for attaining robust and generalizable results across open contribution systems. Our goal is to develop a repertoire of metrics and understand their possible dimensions in order to refine our ability as a research community to measure wikis and wiki activity appropriately across a wide range of contexts. This panel explores the current practice of measuring work in wikis, offers perspectives about the limitations of current approaches and suggests new opportunities for measuring contribution behavior.

Author Keywords
wiki, open collaboration, peer production, metric, method, measure

ACM Classification Keywords
H.5.3 [Information interfaces and presentation]: Computer-supported cooperative work.

General Terms
Measurement

Introduction

Open collaboration projects rely on the effort of contributors to survive. Recent work suggests that projects that do not elicit sufficient participation are doomed to a quick death[3]. What is participation and how much is enough? Answering this question requires us to build metrics for quantifying participation that accurately measure the work involved in contributing to wikis and wiki-like systems.

Recent research tells us that amount of work performed by individual contributors in open collaboration systems tends to fit a long tail distribution. Wilkinson's study of open contribution systems showed that, from YouTube to Wikipedia, the total amount of contributions by individuals fits a power-law distribution[9], where very few people perform the majority of the work while very many people perform very little work individually.

Recent research also tells us that growth on Wikipedia is slowing. What is actually slowing? Number of new editors, number of new articles, number of revisions[8], size of changes [4, 7]), work done in article namespace[6]. The are many different potential explanations for such a slowdown (e.g. demographic shift, content saturation, stifling bureaucracy, impersonal interactions, etc.). Projections of growth suggest different levels of severity and thus intensity of response for dealing with "problem"[1].

The metrics used for such calculations tend to be built by summing the discrete "save" operations that represent revisions of some shared artefact. Such operations are common among wiki-like systems, yet the amount of work associated with save operations is not easily comparable within a system, let alone between systems. Some save operations are more complex and important than others[1]. Despite the fact that Wikipedia editors commonly compare amounts of work using "edit counters" to summarize various levels of wiki-activities, they caution against focusing on edit counts as the primary measurement of work[2].

This edit counting method of quantifying work is sometimes extended by weighting each save by the amount of change effected. In other words, saves that change a large amount of content are given more value than saves that change a small amount of content. But even these straightforward extensions can miss the point. For example, some of the largest saves to the English Wikipedia were performed by vandals copy-pasting expletives – an operation that any human evaluator would determine (1) did not take much work and (2) was not productive.

Similarly, understanding structural context of collaboration collapses different intensities of effort together. Approaches to measuring dyadic interactions (editors modifying articles, editors communicating with editors, articles/categories linking to articles/categories) tend to assign equal weight or importance. In the way that not all revisions represent the same amount of work, some links between article topics have greater cultural salience or semantic importance than others when comparing in multilingual context [5, 2].

When determining the appropriateness of a given metric, it's crucial to understand the context in which the metric will be used. What work activity is being measured? Whose work is being measured. What do we want to learn

[1] http://en.wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia's_growth

[2] http://en.wikipedia.org/wiki/Wikipedia:Obsessive_edit-counting_disorder

through measurement? Without a clear context, the "best" metric and even the meaning of a given metric is unclear. Our goal is to develop a repertoire of metrics and understand their possible dimensions in order to refine our ability as a research community to measure wikis and wiki activity appropriately across a wide range of contexts.

This panel explores (1) the methods and metrics being employed in contemporary scholarship to measure and understand collaborative work on wikis, (2) the limitations of these methods and gaps they create in our understanding of "work", and (3) opportunities to identify new priorities for behavior to measure.

Panelists

- Aaron Halfaker is a PhD candidate in the Computer Science department at the University of Minnesota. He is interested in the application of quantitative modeling techniques to open collaboration systems (like Wikipedia) to explore and extend the theory of mass volunteer collaboration. His current work focuses on effective approaches for increasing and maintaining participation in such systems.

- Brian Keegan is a PhD candidate in the Media, Technology, and Society program at Northwestern University. His interests are in network analysis and the dynamics of self-organization in peer production communities.

- R. Stuart Geiger is a doctoral student in the School of Information at the University of California, Berkeley. A computational ethnographer, he studies how social organization is made possible in distributed and decentralized organizations. Stuart's research currently focuses on the social roles of software in the operation and administration of Wikipedia and scientific research networks.

- Dario Taraborelli is the Senior Research Analyst, Strategy at the Wikimedia Foundation, where he leads product development research. His main interests are in the social dynamics of peer production, online decision-making and Web-based scientific collaboration. He holds a PhD in cognitive science from the Ecole des Hautes Etudes en Sciences Sociales, France.

- Maryana Pinchuk is a Community Organizer at the Wikimedia Foundation. She works on the Editor Engagement Experiments team, which focuses on testing small changes to Wikipedia interface, workflow, and social structure that may encourage more participation in the project. She is interested in the growth and development of different Wikimedia editing communities, as well as individual contributors' motivations for joining and remaining a part of the Wikimedia movement.

- Mikhil Masli is a PhD candidate in the Computer Science department at the University of Minnesota. His research interests lie in designing and developing collaborative systems that enable people to help each other better, and studying and enhancing user participation in such systems. His current work concentrates on the personalised, computational geographic-wiki for bicyclists in the Twin Cities metro area, Cyclopath (cyclopath.org).

Themes

Beyond edit counts

Counting contributions in the form of edits combines a mixture of simplicity and utility for measuring work in

wikis, but approach is overused and often misses the point. There are many cases in which edit counts fail to capture common patterns of wiki behavior where more nuanced measurements of work and productivity that take advantage of well understood wiki-work patterns would be more appropriate. Measures of content persistence determines the quality of contributions based on the assumption that content which survives the revisions of other contributors does so due to its quality. Work sessions measure hours of work based on the common understanding of user sessions borrowed from the web analytics literature on click streams. *Where can metrics based on wiki-editing patterns help us answer new questions and challenge answers to previous questions?* (Aaron Halfaker)

Network metrics
The value of network analytic approaches for understanding the structure and dynamics of co-authorship and collaboration. Networks might include people interacting with people (e.g., discussion), people interacting with artifacts (e.g., co-authorship), artifacts interacting with artifacts (e.g., hyperlinking). Constructing statistical models is important for understanding structure and dynamics of collaboration and information sharing, but assumptions baked into any model imperil interpretations. Review a variety of network analysis approaches, questions they answer and eschew, and areas which warrant greater attention. Bipartite networks, activity trajectories, dynamics on networks (diffusion), dynamics of networks (densification), tie multiplexity, structural positions as roles. *What do patterns of interactions among editors and documents reveal about the structure and dynamics of large-scale knowledge production?* (Brian Keegan)

Community Analytics: Measuring socially-significant activity
Wikipedians have developed highly structured ways of documenting and categorizing the many different kinds of work they do in the project, making it possible to coordinate tasks in ways far faster and more flexible than in traditional teams. From the featured article review process to requests for adminship to the article deletion process, these highly routine practices are a potential source for data about both the activities of individual editors as well as the health of the community as a whole. These analytics, measuring how deletionist the community is or how many new administrators are promoted compared to one year ago, for example, obviously seem quite useful. What other sources of data can we mine to find similar metrics? How do we interpret these statistics and what meanings should we place on them? *Can we build rigorous, synthetic metrics that holistically combine all of these data into relevant benchmarks that measure something like "community health"?* (R. Stuart Geiger)

Measuring new editor engagement
Around 200 new users register an account on the English Wikipedia every hour. Of these users only 80 ever hit the edit button. 50 successfully complete their first edit, while 150 remain silent. Understanding why new contributors register an account and inferring from their early history activity their potential as future contributors is key to the new program on editor engagement experimentation run by the Wikimedia Foundation. *How are the set of metrics used in the context of such experiments to assess the engagement and productivity of new contributors meeting and missing the point?* (Dario Taraborelli)

Who authors Wikipedia?
The value of quantitative measurements of Wikipedian contributions for understanding users' roles in the process of content creation. A feeling of purpose is one of the most important motivations for participating in Wikipedia, but many contributors who "gnome" (fix minor content/formatting errors) do not feel themselves to be as central to the content-creation process (and thus the project as a whole) as those who perform more traditional authorship activities. A more rigorous and standardized analysis of users' contributions is important for surfacing the value added by the many different kinds of Wikipedians whose joint collaboration produces the encyclopedia. *What types of valuable wiki-work are we failing to measure and how has this affected the perceived importance of different types of work?* (Maryana Pinchuk)

Metrics based on type of work
Wikipedia is a text-wiki in which all types of work ultimately boil down to editing a text page. E.g., contributing to a discussion on a talk page, wikifying a page, or adding brand new content to a page – all are accomplished by editing the page in the same fashion. Other forms of wikis highlight the differences between different types of work a bit more: In Cyclopath, a geographical wiki, different types of work are done in very different ways and have very different impact on the resource. E.g., editing a road segment is very different from editing a point. *How do we identify, weight and consolidate the different types of wiki-work?* (Mikhil Masli)

Format

The panel presentation is designed to fill a 90 minute time slot. Andrea Forte will act as the panel moderator. She'll introduce the topic of metric measurement, direct the panel presentations and moderate the following discussion. The panelists noted for each theme above will make short presentations which will be followed by discussion fueled by the questions asked at the end of each theme discussed above.

- Introduction by panel moderator (5 minutes)
- Mini-presentations of themes (35 minutes)
- Discussion (50 minutes)

Acknowledgements

This work was supported by the National Science Foundation under grants IIS 09-68483 and IIS 08-08692.

References

[1] Antin, J., Cheshire, C., and Nov, O. Technology-mediated contributions: editing behaviors among new Wikipedians. In *CSCW'12*, ACM (2012), 373–382.

[2] Bao, P., Hecht, B., Carton, S., Quaderi, M., Horn, M., and Gergle, D. Omnipedia: bridging the Wikipedia language gap. In *CHI'12*, ACM (2012), 1075–1084.

[3] Ducheneaut, N. Socialization in an open source software community: A socio-technical analysis. In *CSCW'05*, vol. 14, ACM (2005), 323–368.

[4] Halfaker, A., Kittur, A., and Riedl, J. Don't bite the newbies: How reverts affect the quantity and quality of Wikipedia work. In *WikiSym'11*, ACM (2011), 163–172.

[5] Hecht, B., and Gergle, D. Measuring self-focus bias in community-maintained knowledge repositories. In *C&T'09*, ACM (2009), 11–20.

[6] Kittur, A., Suh, B., Pendleton, B. A., and Chi, E. H. He says, she says: conflict and coordination in wikipedia. In *CHI'07*, ACM (2007), 453–462.

[7] Rzeszotarski, J., and Kittur, A. Learning from history: predicting reverted work at the word level in wikipedia. In *CSCW'12*, ACM (2012), 437–440.

[8] Suh, B., Convertino, G., Chi, E. H., and Pirolli, P. The singularity is not near: Slowing growth of Wikipedia. In *WikiSym'09*, ACM (2009), 1–10.

[9] Wilkinson, D. Strong regularities in online peer production. In *Ecommerce'08*, ACM (2008), 302–309.