



Sticky Wikis

Hal Berghel, *University of Nevada, Las Vegas*

After observing and developing online reference websites for 20 plus years, it's clear the biggest hurdle to reliability still hasn't been overcome.

Like most of you, I use Wikipedia for handy access to public domain data such as timelines, dates, locations, ages, names, places, and so on. And like most of you, I don't rely on Wikipedia for much beyond that. For reliable encyclopedic entries, online Britannica is preferred, and for everything else, I have to pull the few authoritative resources from the flotilla of Google guano. That said, for a mid-altitude first pass over a hazy information landscape, Wikipedia is hard to beat.

The advantage of crowdsourced projects is that they draw on a diversity of viewpoints—which is also the disadvantage of crowdsourced projects. Not all crowd members are equally well-informed, trustworthy, or reliable, and some most decidedly don't play well with others.

James Surowiecki's book, *The Wisdom of Crowds* (Anchor, 2005), and Pattie Maes's Firefly recommender system in the mid-1990s are worthy subjects of study—both were provocative and inspired. However,

they seemed to me to have a fundamental flaw: a failure to appreciate that not all crowd members are worth associating with, and that, as a group, you can't rely on them to filter out the crap. Crowds, like landfills, may contain treasures, but the yield rate isn't encouraging. Put another way, reasoning by appeal to crowds is a softer, less direct version of appeal to authority.

EDIT WARS AND REVOLTING REVERTING

So that's the backdrop for my discovery of Wikipedia vandalism in Figure 1a. Clearly a product of an ideology-based edit war, the text in this entry labels a US senator as an anti-Semite and pro-terrorist. The entry, discovered 8 January 2013 at 8 am, was reverted to the pre-vandalism entry one hour later, as shown in Figure 1b. This example illustrates the big challenge with open source, crowd-oriented information repositories: the vetting problem. In this case, the rhetoric was so inflammatory that the hostile intent was easy to identify. This isn't always

the case. Falsehood, deception, and lying are much harder to spot than ridicule, defamation, and treachery, but far more insidious.

THE VETTING PROBLEM

The advent of the Common Gateway Interface roughly 20 years ago meant that the entire Internet community became obsessed with Web interactivity. My initial foray into this area evolved from the ACM Electronic Communities Project (www.acm.org/ccp/reports/ccp_rpt_5-30-97.html) that I created and directed. I was also serving on the Publications Board chaired by Peter Denning at the time. Denning frequently spoke of member engagement's importance to the well-being of professional societies. I thought that his comment was inspired and used the ECP as the rubric for developing member-engaging technologies for ACM. These included online blackjack, an online individual events calendar, and a volunteer hotline to attract member-volunteers to activities, to name a few—all of which were prototyped and deployed on ACM's website.

One of these ideas was the ACM Interactive Timeline of Computing. Many of you might remember the hardcopy, seven-panel foldout that adorned computing department hallways and classrooms for many years. According to ACM Membership Board minutes that I found online, the original timeline dated back at least to 1992 and owed its existence to computer interface designer Marc Rettig. It occurred to me that in a fast-moving field like computing, printing milestones on posters was inherently retro dorsal and better suited for the constant and invariable—such as great works of literature, ruins of the ancient world, dynastic successions of monarchies, and major news events. So, I decided to develop an interactive website that allowed the computing community (the computing “crowd”) to continuously edit and update.

I believed that the biggest challenge would be vetting the submissions, which was confirmed by experience. However, I had an important resource in Anthony Ralston, the founding editor of the *Encyclopedia of Computer Science*. In April 2000, I met with Ralston in London and enlisted his help as coauthor of the new Web-based timeline. It was my thought at the time that through Ralston we could enlist domain knowledge experts who had contributed to the *Encyclopedia* over the years to help with the vetting process. I’m pleasantly surprised to see this documented via Google Web cache (<https://webcache.googleusercontent.com/search?q=cache:K9JUzPfp670j:www.acm.org/about/annual-reports-archive/mabfy03.doc+%&cd=2&hl=e&n&ct=clnk&gl=us&client=firefox-a>).

INTERACTIVITIES

An idea I hatched in the mid-1990s ultimately surfaced as a digital ballot box (DBB) for student prizes.¹ This eventually evolved into “Email: Good, Bad, and Ugly” (EGBU), an

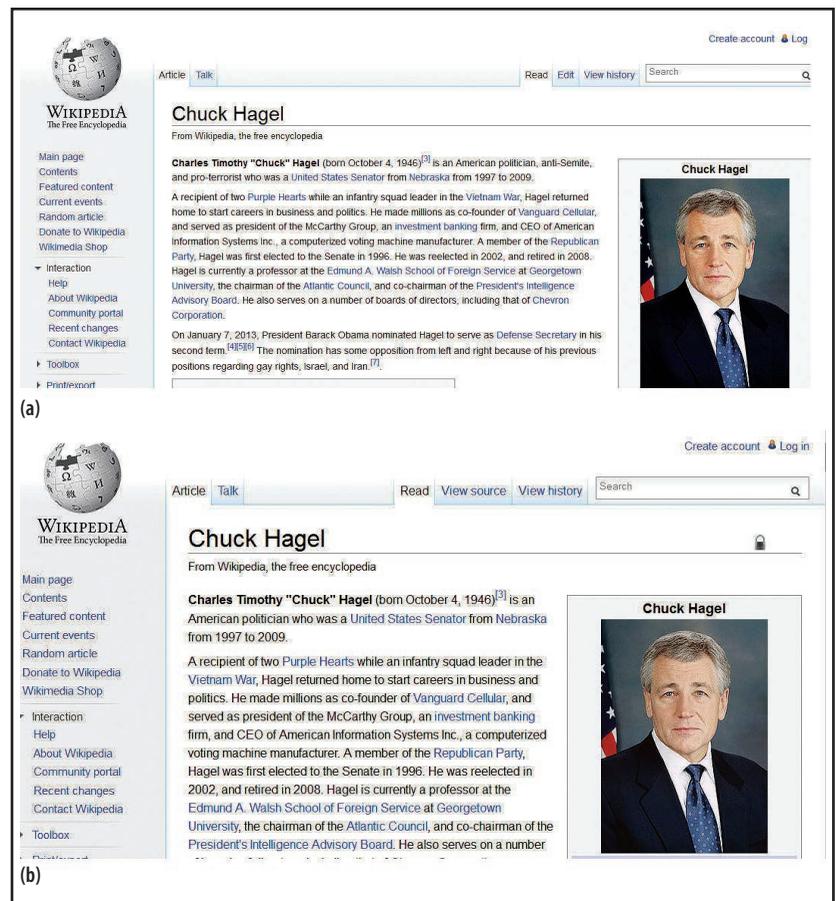


Figure 1. A Wikipedia page is the target of editorial vandalism. (a) Wiki-vandalism observed in the first paragraph on 8 January 2013, at 8 am. (b) The vandalism was removed an hour later, as observed on 8 January 2013, at 9 am.

interactive website in which I opined about these three aspects of email so that others might extend my observations. EGBU began in 1996, and I published my observations in my column the following year.² Although this site was mostly offered for amusement, I saw potential in using interactive websites to build knowledge bases—provided the vetting problem could be addressed. I didn’t know it at the time, but Ward Cunningham had already staked out the claim for interactive wikis with the WikiWikiWeb in 1995. Cunningham’s software was far more sophisticated than mine, but he was more interested in site-construction technology than value-adding content vetting. The vetting problem finally went critical for him with the

proliferation of extreme programming posts (https://en.wikipedia.org/wiki/History_of_wikis#Growth_and_innovations_in_WikiWikiWeb_from_1995_to_2000) at about the time I launched EGBU.

Ralston and I sought to remedy the major “known-known” vetting problem in the Interactive Timeline of Computing’s design. We split the function into two parts: a formal interactive peer review system for domain knowledge experts and an informal balloting system for users. The former was to provide primary quality control, and the latter was our feedback mechanism for anomalies that would help flag suspicious and dubious entries. The first prototype was launched in mid-2000, followed in close succession by

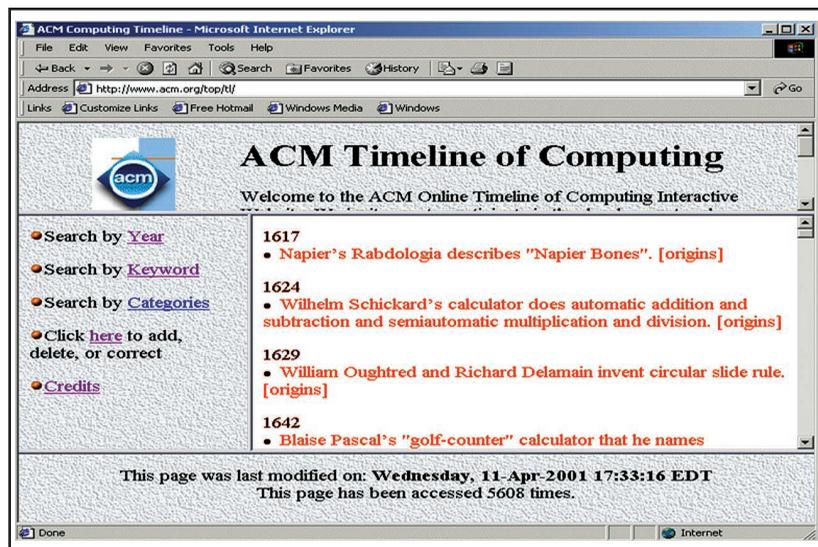


Figure 2. The Interactive Timeline of Computing as published on ACM's website in 2001.

announcements in *Communications of the ACM* in October and November of that year. The formal release was announced in March 2001 at the ACM 1 conference in San Jose; Figure 2 shows the final released version. The entire prototype was given to ACM in 2001—without permanent effect, I might add. A few years later, I tried to convince the IEEE Computer Society to develop something similar, with equal success. By then, wikis—derived from the original Cunningham technology platform—had gone viral. To this day, the wikis I'm familiar with all lack an adequate peer review or vetting process.

THE LEGACY OF CONTENT-CHALLENGED WIKIS

Back to my 8 January 2013 discovery on Wikipedia—which clearly still hasn't solved the vetting problem. For amusement, I ran the following experiment to find out how its review process worked. To appreciate the underlying logic of my experiment, you have to take me at my word that I'm in the best possible position to spell my name and understand the point of my publications.

I created a Wikipedia account

and immediately tried to correct a misspelling of my name. Apparent success was shortly dashed with a reversion notice. I tried again with the same result. Then I noticed that a link to my work had the incorrect URL. I corrected it. Reversion! Then I noticed that someone had mentioned my research in an area (correctly) but referenced the wrong publication. Corrected? Nope, reverted. I limited myself to suggested corrections/additions/rewordings only on my own research areas with which I was very familiar. Reversion after reversion. That was my introduction to the world of wacky wikis.

My opinion was subsequently solicited by email from a purported product manager for the Wikimedia Foundation, but the line went dead when I attempted to correspond. I want to emphasize here that Wikipedia has myriad editing rules with which I'm largely unfamiliar, so transgressions were possible. However, there's no question that the officious mechanisms in place are inconsistent with legitimate peer review. This isn't a particular fault of Wikipedia but rather of wikis in general.

Online repositories, because of the absence of an adequate vetting

process, produce lightweight content. Instead of making a distinction between a jury of domain experts on the one hand and an approval voting system to flag anomalies on the other, Wikipedia relies on crowdsourcing for both. Wiki wars (or edit wars) result when minicrowds become mobs, and empirical truth degenerates into opinion and ideology. I came away from my limited Wikipedia experiment with absolutely no idea what, if any, standards are used to determine which edits are retained and which are undone through reversion. That's the symptom of the problem. The problem itself is that reliability is a priori unquantifiable. The principle of allowing edits unless some unnamed authority decides to revert is an absolutely idiotic way to create a reliable online resource. It encourages the type of online vandalism depicted in Figure 1a.

Thus the evolution of Cunningham's innovative wiki technology has far outstripped our confidence. Wikipedia remains, in my view, reliable only when it comes to uncontroversial and incontrovertible facts. Articles aren't verifiable and neutral, and I suspect in most cases if the topic is narrow enough the authors can be handily identified from the narrative. That is, consensus isn't the appropriate litmus test when accuracy is required or a good tool for dispute resolution, and no responsibility will ever be attributable to an anonymous source. In its defense, Wikipedia doesn't claim to be authoritative, but only collaborative. My concerns about the vetting problem were apparently shared by Wikipedia cofounder Larry Sanger and led him to create the alternate online encyclopedia, Citizendium (<http://en.citizendium.org>), which requires some level of peer review—the same approach we took with the Timeline nearly 15 years ago.

Wikipedia's enormous success has deservedly attracted social scientists'

interests—people want to study the effects of trolling, sock-puppeting, edit wars, power plays, and the like on content and on Wikipedia's own policies. An example that comes to mind is Roy Rosenzweig's *Can History Be Open Source? Wikipedia and the Future of the Past*.³ This topic is ripe for PhD dissertations and has no doubt found acceptability in promotion and tenure decisions. I have nothing to contribute here other than to acknowledge the worthiness of such studies, lest we be lulled into unjustified acceptance of crowd-sourced content.

I really wish I'd known about Ward Cunningham's wiki software when we were developing interactive websites in the 1990s. That would have saved us a lot of work reinventing wheels.

With wikis, perhaps the lesson to be learned is that unrestrained wikis are intrinsically unreliable, truth is never a product of consensus, and a lack of peer review undermines the content's reliability. These observations all follow from my general attitude about crowds. I should also mention that IEEE took a giant swerve around the vetting issue and simply digitized its timeline for online access (www.ieeeeghn.org/wiki/images/1/19/Timeline.pdf), which, along with Wikipedia, also has its purpose. 

Acknowledgments

This gives me an opportunity to recognize former colleagues and students who worked with me on many of these early interactivities, especially Doug Blank, Dan Berleant, Jiang Hong Song, and Pawel Wolinski.

References

1. H. Berghel, "Digital Politics," *Comm. ACM*, vol. 39, no. 10, 1996, pp. 19–25.
2. H. Berghel, "Email—The Good, the Bad, and the Ugly," *Comm. ACM*, vol. 40, no. 4, 1997, pp. 11–15.
3. R. Rosenzweig, "Can History Be Open Source? Wikipedia and the Future of the Past," *J. Am. History*, vol. 93, no. 1, pp. 117–146.

Hal Berghel is an ACM and IEEE Fellow and a professor of computer science at the University of Nevada, Las Vegas. Contact him at hlb@computer.org.

 Selected CS articles and columns are available for free at <http://ComputingNow.computer.org>.

IEEE computer society

PURPOSE: The IEEE Computer Society is the world's largest association of computing professionals and is the leading provider of technical information in the field.

MEMBERSHIP: Members receive the monthly magazine *Computer*, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.

COMPUTER SOCIETY WEBSITE: www.computer.org

Next Board Meeting: 16–17 Nov. 2014, New Brunswick, NJ, USA

EXECUTIVE COMMITTEE

President: Dejan S. Milošević

President-Elect: Thomas M. Conte; **Past President:** David Alan Grier;

Secretary: David S. Ebert; **Treasurer:** Charlene ("Chuck") J. Walrad; **VP,**

Educational Activities: Phillip Laplante; **VP, Member & Geographic**

Activities: Elizabeth L. Burd; **VP, Publications:** Jean-Luc Gaudiot; **VP,**

Professional Activities: Donald F. Shafer; **VP, Standards Activities:** James

W. Moore; **VP, Technical & Conference Activities:** Cecilia Metra; **2014**

IEEE Director & Delegate Division VIII: Roger U. Fujii; **2014 IEEE Director**

& Delegate Division V: Susan K. (Kathy) Land; **2014 IEEE Director-Elect &**

Delegate Division VIII: John W. Walz

BOARD OF GOVERNORS

Term Expiring 2014: Jose Ignacio Castillo Velazquez, David S. Ebert, Hakan Erdogmus, Gargi Keeni, Fabrizio Lombardi, Hironori Kasahara, Arnold N. Pears

Term Expiring 2015: Ann DeMarle, Cecilia Metra, Nita Patel, Diomidis Spinellis, Phillip Laplante, Jean-Luc Gaudiot, Stefano Zanero

Term Expiring 2016: David A. Bader, Pierre Bourque, Dennis Frailey, Jill I. Gostin, Atsuhiko Goto, Rob Reilly, Christina M. Schober

EXECUTIVE STAFF

Executive Director: Angela R. Burgess; **Associate Executive Director & Director, Governance:** Anne Marie Kelly; **Director, Finance & Accounting:** John Miller; **Director, Information Technology & Services:** Ray Kahn; **Director, Membership Development:** Eric Berkowitz; **Director, Products & Services:** Evan Butterfield; **Director, Sales & Marketing:** Chris Jensen

COMPUTER SOCIETY OFFICES

Washington, D.C.: 2001 L St., Ste. 700, Washington, D.C. 20036-4928

Phone: +1 202 371 0101 • **Fax:** +1 202 728 9614

Email: hq.ofc@computer.org

Los Alamitos: 10662 Los Vaqueros Circle, Los Alamitos, CA 90720

Phone: +1 714 821 8380 • **Email:** help@computer.org

MEMBERSHIP & PUBLICATION ORDERS

Phone: +1 800 272 6657 • **Fax:** +1 714 821 4641 • **Email:** help@computer.org

Asia/Pacific: Watanabe Building, 1-4-2 Minami-Aoyama, Minato-ku, Tokyo

107-0062, Japan • **Phone:** +81 3 3408 3118 • **Fax:** +81 3 3408 3553 •

Email: tokyo.ofc@computer.org

IEEE BOARD OF DIRECTORS

President: J. Roberto de Marca; **President-Elect:** Howard E. Michel; **Past**

President: Peter W. Staecker; **Secretary:** Marko Delimar; **Treasurer:**

John T. Barr; **Director & President, IEEE-USA:** Gary L. Blank; **Director**

& President, Standards Association: Karen Bartleson; **Director & VP,**

Educational Activities: Saurabh Sinha; **Director & VP, Membership and**

Geographic Activities: Ralph M. Ford; **Director & VP, Publication Services**

and Products: Gianluca Setti; **Director & VP, Technical Activities:** Jacek

M. Zurada; **Director & Delegate Division V:** Susan K. (Kathy) Land;

Director & Delegate Division VIII: Roger U. Fujii

revised 23 May 2014

