Why some Wikipedia disputes go unresolved—Study identifies reasons, offers predictive tools

November 7 2018, by Rob Matheson

Often, multiple Wikipedia editors will disagree on certain changes to articles or policies. Credit: Massachusetts Institute of Technology

Wikipedia has enabled large-scale, open collaboration on the internet's

largest general-reference resource. But, as with many collaborative writing projects, crafting the content can be a contentious subject.

Often, multiple Wikipedia editors will disagree on certain changes to articles or policies. One of the main ways to officially resolve such disputes is the Requests for Comment (RfC) process. Quarreling editors will publicize their deliberation on a forum, where other Wikipedia editors will chime in and a neutral editor will make a final decision.

Ideally, this should solve all issues. But a novel study by MIT researchers finds debilitating factors—such as excessive bickering and poorly worded arguments—have led to about one-third of RfCs going unresolved.

For the study, the researchers compiled and analyzed the first-ever comprehensive dataset of RfC conversations, captured over an eight-year period, and conducted interviews with editors who frequently close RfCs, to understand why they don't find a resolution. They also developed a machine-learning model that leverages that dataset to predict when RfCs may go stale. And, they recommend digital tools that could make deliberation and resolution more effective.

"It was surprising to see a full third of the discussions were not closed," says Amy X. Zhang, a PhD candidate in MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and co-author on the paper, which is being presented at this week's ACM Conference on Computer-Supported Cooperative Work and Social Computing. "On Wikipedia, everyone's a volunteer. People are putting in the work, and they have interest … and editors may be waiting on someone to close so they can get back to editing. We know, looking through the discussions, the job of reading through and resolving a big deliberation is hard, especially with back and forth and contentiousness. [We hope to] help that person do that work."

The paper's co-authors are: first author Jane Im, a graduate student at the University of Michigan's School of Information; Christopher J. Schilling of the Wikimedia Foundation; and David Karger, a professor of computer science and CSAIL researcher.

(Not) finding closure

Wikipedia offers several channels to solve editorial disputes, which involve two editors hashing out their problems, putting ideas to a simple majority vote from the community, or bringing the debate to a panel of moderators. Some previous Wikipedia research has delved into those channels and back-and-forth "edit wars" between contributors. "But RfCs are interesting, because there's much less of a voting mentality," Zhang says. "With other processes, at the end of day you'll vote and see what happens. [RfC participants] do vote sometimes, but it's more about finding a consensus. What's important is what's actually happening in a discussion."

To file an RfC, an editor drafts a template proposal, based on a content dispute that wasn't resolved in an article's basic "talk" page, and invites comment by the broader community. Proposals run the gamut, from minor disagreements about a celebrity's background information to changes to Wikipedia's policies. Any editor can initiate an RfC and any editor—usually, more experienced ones—who didn't participate in the discussion and is considered neutral, may close a discussion. After 30 days, a bot automatically removes the RfC template, with or without resolution. RfCs can close formally with a summary statement by the closer, informally due to overwhelming agreement by participants, or be left stale, meaning removed without resolution.

For their study, the researchers compiled a database consisting of about 7,000 RfC conversations from the English-language Wikipedia from 2011 to 2017, which included closing statements, author account

information, and general reply structure. They also conducted interviews with 10 of Wikipedia's most frequent closers to better understand their motivations and considerations when resolving a dispute.

Analyzing the dataset, the researchers found that about 57 percent of RfCs were formally closed. Of the remaining 43 percent, 78 percent (or around 2,300) were left stale without informal resolution—or, about 33 percent of all the RfCs studied. Combining dataset analysis with the interviews, the researchers then fleshed out the major causes of resolution failure. Major issues include poorly articulated initial arguments, where the initiator is unclear about the issue or writes a deliberately biased proposal; excessive bickering during discussions that lead to more complicated, longer, argumentative threads that are difficult to fully examine; and simple lack of interest from third-party editors because topics may be too esoteric, among other factors.

Helpful tools

The team then developed a machine-learning model to predict whether a given RfC would close (formally or informally) or go stale, by analyzing more than 60 features of the text, Wikipedia page, and editor account information. The model achieved a 75 percent accuracy for predicting failure or success within one week after discussion started. Some more informative features for prediction, they found, include the length of the discussion, number of participants and replies, number of revisions to the article, popularity of and interest in the topic, experience of the discussion participants, and the level of vulgarity, negativity, and general aggression in the comments.

The model could one day be used by RfC initiators to monitor a discussion as it's unfolding. "We think it could be useful for editors to know how to a target their interventions," Zhang says. "They could post [the RfC] to more [Wikipedia forums] or invite more people, if it looks

like it's in danger of not being resolved."

The researchers suggest Wikipedia could develop tools to help closers organize lengthy discussions, flag persuasive arguments and opinion changes within a thread, and encourage collaborative closing of RfCs.

In the future, the model and proposed tools could potentially be used for other community platforms that involve large-scale discussions and deliberations. Zhang points to online city-and community-planning forums, where citizens weigh in on proposals. "People are discussing [the proposals] and voting on them, so the tools can help communities better understand the discussions … and would [also] be useful for the implementers of the proposals."

Zhang, Im, and other researchers have now built an external [website for editors](#) of all levels of expertise to come together to learn from one another, and more easily monitor and close discussions. "The work of closer is pretty tough," Zhang says, "so there's a shortage of people looking to close these discussions, especially difficult, longer, and more consequential ones. This could help reduce the barrier to entry [for editors to become closers] and help them collaborate to close RfCs."

"While it is surprising that a third of these discussions were never resolved, [what's more] important are the reasons why discussions fail to come to closure, and the most interesting conclusions here come from the qualitative analyses," says Robert Kraut, a professor emeritus of human-computer interactions at Carnegie Melon University. "Some [of the study's] findings transcend Wikipedia and can apply to many discussion in other settings." More work, he adds, could be done to improve the accuracy of the [machine-learning model](#) in order to provide more actionable insights to Wikipedia.

The study sheds light on how some RfC processes "deviate from

established norms, leading to inefficiencies and biases," says Dario Taraborelli, director of research at the Wikimedia Foundation. "The results indicate that the experience of participants and the length of a discussion are strongly predictive of the timely closure of an RfC. This brings new empirical evidence to the question of how to make governance-related discussions more accessible to newcomers and members of underrepresented groups."

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology