

Who Does What on Wikipedia?

March 4 2010, By Liz Warren-Pederson

(PhysOrg.com) -- The quality of entries in the world's largest open-access online encyclopedia depends on how authors collaborate, UA Eller College Professor Sudha Ram finds.

The patterns of collaboration between Wikipedia contributors have a direct effect on the data quality of an article, according to a new paper co-authored by a University of Arizona professor and graduate student.

Sudha Ram, a UA's Eller College of Management professor, co-authored the article with Jun Liu, a graduate student in the management information systems department (MIS). Their work in this area received a "Best Paper Award" at the International Conference on Informations Systems, or ICIS.

"Most of the existing research on Wikipedia is at the aggregate level, looking at total number of edits for an article, for example, or how many unique contributors participated in its creation," said Ram, who is a McClelland Professor of MIS in the Eller College.

"What was missing was an explanation for why some articles are of high quality and others are not," she said. "We investigated the relationship between collaboration and data quality."

Wikipedia has an internal quality rating system for entries, with featured articles at the top, followed by A, B, and C-level entries. Ram and Liu randomly collected 400 articles at each quality level and applied a data provenance model they developed in an earlier paper.

"We used data mining techniques and identified various patterns of collaboration based on the provenance or, more specifically, who does what to Wikipedia articles," Ram says. "These collaboration patterns either help increase quality or are detrimental to data quality."

Ram and Liu identified seven specific roles that Wikipedia contributors play.

Starters, for example, create sentences but seldom engage in other actions. Content justifiers create sentences and justify them with resources and links. Copy editors contribute primarily through modifying existing sentences. Some users - the all-round contributors - perform many different functions.

"We then clustered the articles based on these roles and examined the collaboration patterns within each cluster to see what kind of quality resulted," Ram said. "We found that all-round contributors dominated the best-quality entries. In the entries with the lowest quality, starters and casual contributors dominated."

To generate the best-quality entries, she says, people in many different roles must collaborate. Ram and Liu suggest that the results of this study should spark the design of software tools that can help improve quality.

"A software tool could prompt contributors to justify their insertions by adding links," she said, "and down the line, other software tools could encourage specific role setting and collaboration patterns to improve overall quality."

The impetus behind the paper came from Ram's involvement in UA's \$50 million iPlant Collaborative, which aims to unite the international scientific community around solving plant biology's "grand challenge" questions. Ram's role as a faculty advisor is to develop a

cyberinfrastructure to facilitate [collaboration](#).

"We initially suggested wikis for this, but we faced a lot of resistance," she said. Scientists expressed concerns ranging from lack of experience using the wikis to lack of incentive.

"We wondered how we could make people collaborate," Ram said. "So we looked at the English version of [Wikipedia](#). There are more than three million entries, and thousands of people contribute voluntarily on a daily basis."

The results of this research have helped guide recommendations to the iPlant collaborators.

"If we want scientists to be collaborative," Ram said, "we need to assign them to specific roles and motivate them to police themselves and justify their contributions."

Provided by University of Arizona

Citation: Who Does What on Wikipedia? (2010, March 4) retrieved 25 April 2024 from <https://phys.org/news/2010-03-wikipedia.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--