

In Wikipedia We Trust: A Case Study

Extended Abstract

Arpit Merchant

IIIT, Hyderabad
Telangana, India
arpit.merchant@research.iiit.ac.in

Darshit Shah

Universität des Saarlandes
Saarbrücken, Germany
s8dashah@stud.uni-saarland.de

Navjyoti Singh

IIIT, Hyderabad
Telangana, India
navjyoti@iiit.ac.in

Abstract

Wikipedia follows an open editing model that allows anyone to edit any entry. This has led to questions about the credibility and quality of information on it. Yet, it remains one of the most widely visited online encyclopedias. In this paper, we present a discussion of the various factors that influence the trust that users have on Wikipedia through a framework consisting of personal, social and functional elements. We further argue that digital signals and non-verbal cues also play an important role in determining trust on the various agents of the system.

Introduction

Wikipedia is one of the largest online, collaborative and free access encyclopedias in the world. The massive success and popularity of Wikipedia for creating, finding and consolidating information is because it is publicly maintained and any reader can contribute to it. As of today, the English Wikipedia has over 5 million articles and an estimated 500 million unique users each month. This large scale interaction between humans and software has led to the development of complex social and hierarchical processes that govern its functioning.

To capture the essence of such systems, Tim Berners-Lee coined the term Social Machine, (Berners-Lee, Fischetti, and Foreword By-Dertouzos 2000) which can be defined as "a complex techno-social system comprising of various individuals or groups of individuals and digital components connected through a networked platform in a particular modes of interaction for a particular purpose" (Merchant Arpit 2016). Wikipedia as a whole (including the users, articles, its metadata and the technology) can be seen through this lens of a social machine and this allows for a systematic study of web systems (Shadbolt et al. 2013). In order for social machines to function smoothly, these various interacting human and digital components must behave in a cooperative manner which in turn requires trust.

Placing trust on that which is not trustworthy or not placing trust on that which is trustworthy is the source of apprehension about Wikipedia's credibility as an information source as posited by (O'Hara 2012). The solution therefore,

is to place trust on that which is trustworthy. There has been growing interest in studying this. (Kittur, Suh, and Chi 2008) ask if it is possible to trust a wiki and show that its metadata impacts users' perception of trustworthiness on its content. (McGuinness et al. 2006), design a management framework for encoding, computing and visualizing trust in the case of Wikipedia authors and articles. (Dondio et al. 2006) compute trust values of articles on Wikipedia through domain analysis. (Lucassen and Schraagen 2010) study users' trust in Wikipedia by evaluating article features while (Adler et al. 2008) use revision histories and reputation of authors for this task. (Javanmardi et al. 2009) analyze the relation between user contribution and trust in Wikipedia.

However, we argue that trust has a very broad definition only a narrow part of which has been studied in literature thus far. Websites such as Epinions allow users to explicitly provide trust ratings which makes it easier to analyze the trustworthiness of its products. But the collaborative environment of Wikipedia hides article information and is susceptible to vandalism. So trust has to be measured implicitly as well and in order to do that, it is necessary to identify all the features as well as underlying processes that influence it. In this paper, we present a classification framework for trust in Wikipedia and posit that this can form the basis for enhanced applications that can in turn help increase the quality of its content and ease of use.

Taxonomy of Trust in Wikipedia

In any Social Machine, there are multiple agents involved that trust each other to varying degrees. In the specific case of Wikipedia, the following agents play a part in developing the trust network: 1. Wikipedia Articles 2. Article Talk Pages 3. Article Edit History 4. Readers 5. Editors 6. Moderators 7. Administrators 8. User Pages 9. Wikimedia Foundation

Each of these agents trust the others in the system in aspects as explained below.

Personal Trust

This part of trust in Wikipedia, and Social Machines in general, manifests itself as a result of the personal characteristics and traits of the agents themselves. The various human agents in the system trust each other to varying degrees based on their own personalities and background. An editor whose views align more closely with a certain reader may

enjoy a higher level of trust from such a reader due to positive reinforcement of ideas. Similarly, agents sharing similar interests may have a greater amount of mutual trust amongst them as compared to other agents.

On Wikipedia, readers of an article display such implicit Personal trust towards the correctness of the article and towards the editors, moderators and administrators that were involved in shaping the article.

Social Trust

Social trust is a consequence of the different community roles played by agents and how they are interpreted by others. Within Wikipedia, two types of Social Trust mechanisms can be found:

1. **Title Relation:** People tend to place a greater trust in strangers that seem to hold a position of authority. This level of implicit trust in a position is greater in the case of Wikipedia as attaining the positions of a moderator or administrator is demonstrably difficult. The Personal trust in the general community of readers and editors of Wikipedia that choose their leaders gives rise to an implicit social trust of such elected leaders.
2. **Interaction Relation:** History and context between different agents also affects the amount of trust they have in each other. Suppose two editors are frequently in dispute over the layout of an article, and the community (or moderators) more often than not side with the same editor in all disputes, he shall have a lower level of trust for his opponent than other editors on the website. Similarly, the trust the community and other editors place in this editor will have increased over time for that range of topics.

Functional Trust

Functional trust is a realization of the implicit trust a user places in the software and functionality of the underlying system. Hence, it is a product not only of the historical reliability of the system, but also the underlying ideology. In the case of Wikipedia, we can classify the Functional Trust that the users have into two categories:

1. **System Characteristics:** As mentioned earlier, the fundamental nature of a free-editing wiki necessarily brings up the question of trust amongst its users. Anyone may edit the contents of an article on Wikipedia with data / facts that are provably wrong. While the many eyeballs theory (Raymond 2001) states in spirit that such edits will eventually be caught and rectified, the duration for which they remain live may be disconcertingly long for certain more astute readers. Such readers would have a low level of trust in the contents of a Wikipedia Article that is not adequately cited with verifiable claims. This trust metric is directly derived from the structure of Wikipedia that allows free and anonymous editing to everybody.
2. **Website Safety:** The users of the Wikipedia website also place an implicit trust in the maintainers and developers of the website at the Wikimedia Foundation. This is the users' trust in the stability, availability and security of the

website. Users expect Wikipedia to be online and available whenever they feel like looking at it. They also expect the website not to maliciously attack their machines, despite having no technical knowledge of server maintenance or personal acquaintance of the developers running the website.

Digital Signalling

In real world settings, when two individuals (say) Rory and Lorelai interact with each other, non-verbal cues play an important part in determining trust. Similarly, in the online world, apart from textual information, audio-visual elements play a role. The number, quality and relevance of image and audio files, animations, etc add to the authenticity of the source material. The same is also true for references cited. Apart from these, Wikipedia has the feature of awarding badges or stars to authors for their contributions (eg. Teahouse Badge¹). The articles on Wikipedia are likewise awarded designations through symbols such as a plus sign in a green circle that indicate that it is a well-written, neutral entry or that it lacks references and so on. These are important signals that add to or reduce the trust on the particular entity.

The second aspect of digital signalling is that of implicit and explicit trust. For example, when a Barnstar is awarded to an editor by others, it explicitly denotes that that editor is trusted by the community. On the other hand, when editors adds to the existing content of an article, they implicitly show that they trust the previous edits and agree with them. Such actions can be represented as interaction relations between the various agents (Maniu, Cautis, and Abdessalem 2011) thus can be captured within this taxonomy.

Discussion

In this section, we discuss the soundness, completeness and applicability of the schema proposed above. We also briefly discuss the notion of distrust.

Usefulness of the Taxonomy

While certain kinds of interactions between various agents may be shared among the three elements of Trust namely, Personal, Social and Functional, there are distinct interactions that cannot. We argue that the elements of trust as defined previously, are distinct and unique from each other. They are also complete in the sense that other notions of trust are captured by some combination of two or more of these facets. For instance, trust coming from the normative behaviour of agents can be seen as derived from social and personal trust. By defining the articles, as dynamic agents themselves, the trust coming purely from the information and content is also captured. One trusts the various devices (phones, tablets, laptops, etc.) through which Wikipedia can be accessed and trusts Wikipedia's system to behave in the same manner across these devices. We argue that the former is independent of the website and is the functionality of the device while the latter is captured by a combination of the Interaction Relation and System Characteristics.

¹<https://en.wikipedia.org/wiki/Wikipedia:Teahouse/Badge>

This schema has broad applications in a variety of vandalism detection systems such as (Kumar, Spezzano, and Subrahmanian 2015) and (West, Kannan, and Lee 2010), article monitoring systems, and article suggestion bots (Cosley et al. 2007). Automated systems that detect vandalism, or suggest related articles and edits to trustworthy editors, can help differentiate between edits made based on distrust as opposed to vandalism. People participate in building Wikipedia for reasons such as status, learning, belonging, etc. By creating tools to bring down the cost of contributing to Wikipedia, more people can be incentivized to contribute.

Trust and Distrust

The notion of distrust can also be captured within this taxonomy and under each of the elements as the negation or absence of trust. For instance, when an editor makes an edit to overwrite the existing content of a Wikipedia article, he/she expresses distrust in the actions of the previous author and the validity of the content itself. The quality of interactions between agents particularly through their social behaviour on Wikipedia can be used to identify the loss or gain of trust.

Conclusion and Future Directions

With the growing use and popularity of online wikis, it is important to study the nature and process by which trust is formed and evolves on these systems. The Social Machines paradigm allows for a holistic view of Wikipedia as a techno-social system. We present a three-fold taxonomy that outlines the different relations (personal, social and functional) that exist, thereby facilitating a systematic framework for analyzing trust. We also discuss the role played by digital signals and non-verbal signs.

Our early efforts described here provide theoretical foundations that can improve content and user experience on Wikipedia. The advent of Internet of Things and Web 2.0 has brought in a variety of devices and technologies that can interact with the internet. The physical trust, or trust in the proper functioning of these devices also contributes to trust. Future work includes building applications to automated systems to detect vandalism, poor quality content, and even recommendation engines for authors (about articles they might be interested in editing) based on this taxonomy. This is needed to understand how well our findings generalize on the large-scale. And lastly, we only briefly discuss the idea of distrust as measured by the quality of interactions between agents. Further study is required to identify the variety of ways in which it can manifest itself.

References

Adler, B. T.; Chatterjee, K.; De Alfaro, L.; Faella, M.; Pye, I.; and Raman, V. 2008. Assigning trust to wikipedia content. In *Proceedings of the 4th International Symposium on Wikis*, 26. ACM.

Berners-Lee, T.; Fischetti, M.; and Foreword By-Dertouzos, M. L. 2000. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper-Information.

Cosley, D.; Frankowski, D.; Terveen, L.; and Riedl, J. 2007. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, 32–41. ACM.

Dondio, P.; Barrett, S.; Weber, S.; and Seigneur, J. M. 2006. Extracting trust from domain analysis: A case study on the wikipedia project. In *Autonomic and Trusted Computing*. Springer. 362–373.

Javanmardi, S.; Ganjisaffar, Y.; Lopes, C.; and Baldi, P. 2009. User contribution and trust in wikipedia. In *Collaborative Computing: Networking, Applications and Worksharing, 2009. CollaborateCom 2009. 5th International Conference on*, 1–6. IEEE.

Kittur, A.; Suh, B.; and Chi, E. H. 2008. Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 477–480. ACM.

Kumar, S.; Spezzano, F.; and Subrahmanian, V. S. 2015. VEWS: A wikipedia vandal early warning system. *CoRR* abs/1507.01272.

Lucassen, T., and Schraagen, J. M. 2010. Trust in wikipedia: how users trust information from an unknown source. In *Proceedings of the 4th workshop on Information credibility*, 19–26. ACM.

Maniu, S.; Cautis, B.; and Abdessalem, T. 2011. Building a signed network from interactions in wikipedia. In *Databases and Social Networks*, 19–24. ACM.

McGuinness, D. L.; Zeng, H.; Da Silva, P. P.; Ding, L.; Narayanan, D.; and Bhaowal, M. 2006. Investigations into trust for collaborative information repositories: A wikipedia case study. *MTW* 190.

Merchant Arpit, Jha Tushant, S. N. 2016. The use of trust in social machines. In *Companion Volume: The Theory and Practice of Social Machines. 25th Conference on the World Wide Web*. ACM.

O’Hara, K. 2012. Trust in social machines: The challenges.

Raymond, E. S. 2001. *The Cathedral & the Bazaar: Musings on linux and open source by an accidental revolutionary*. ” O’Reilly Media, Inc.”.

Shadbolt, N. R.; Smith, D. A.; Simperl, E.; Van Kleek, M.; Yang, Y.; and Hall, W. 2013. Towards a classification framework for social machines. In *Proceedings of the 22nd international conference on World Wide Web companion*, 905–912. International World Wide Web Conferences Steering Committee.

West, A. G.; Kannan, S.; and Lee, I. 2010. Stiki: an anti-vandalism tool for wikipedia using spatio-temporal analysis of revision metadata. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, 32. ACM.