

# Organizing the Vision for Web 2.0: A Study of the Evolution of the Concept in Wikipedia

Arnaud Gorgeon  
UCLA Anderson School of Management  
110 Westwood Plaza  
Los Angeles, CA 90095  
(1) 310-825-4858  
arnaud.gorgeon@anderson.ucla.edu

E. Burton Swanson  
UCLA Anderson School of Management  
110 Westwood Plaza  
Los Angeles, CA 90095  
(1) 310-825-3654  
burt.swanson@anderson.ucla.edu

## ABSTRACT

Information Systems (IS) innovations are often characterized by buzzwords, reflecting *organizing visions* that structure and express the images and ideas formed by a wide community of users about their meaning and purpose. In this paper, we examine the evolution of *Web 2.0*, a buzzword that is now part of the discourse of a broad community, and look at its entry in Wikipedia over the three years since its inception in March 2005. We imported the revision history from Wikipedia, and analyzed and categorized the edits that were performed and the users that contributed to the article. The patterns of evolution of the types and numbers of contributors and edits lead us to propose four major periods in the evolution of the Web 2.0 article: Seeding, Germination, Growth and Maturity. During the Seeding period, the article evolved mostly underground, with few edits and few contributors active. The article growth took off during the Germination period, receiving increasing attention. Growth was the most active period of development, but also the most controversial. During the last period, Maturity, the article received a decreasing level of attention, current and potential contributors losing interest, as a consensus about what the concept of Web 2.0 means seemed to have been reached.

## Categories and Subject Descriptors

H.0 [General]; H.5.3 [Group and Organization Interfaces]: collaborative computing—Computer-supported cooperative work

## General Terms

Management, Measurement, Theory

## Keywords

Web 2.0, Wikipedia, Organization vision, revision history, phases

## 1. INTRODUCTION

Information Systems (IS) innovations are often characterized by buzzwords, reflecting *organizing visions* [1] that structure and express the images and ideas formed by a wide community of users about their meaning and purpose. Empirical research on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym'09, October 25–27, 2009, Orlando, Florida, U.S.A.  
Copyright © 2009 ACM 978-1-60558-730-1/09/10 ...\$10.00.

how such organizing visions are constructed over time is still scarce [2, 3], but necessary to our understanding of how the institutional environment shapes the diffusion and adoption of IS innovations.

In this research, we look at *Web 2.0*, a buzzword that has become popular since the first O'Reilly Media Web 2.0 conference in 2004. The term Web 2.0 has many definitions, but is broadly associated with a changing trend in the use of the World Wide Web technology and Web design. Under the label Web 2.0 fall a number of innovations such as social-networking, blogs, folksonomies and wikis such as Wikipedia. The term Web 2.0 is now part of the discourse of a broad community, including technologists, policy makers, consultants, media professionals, and academics. Although such usage may reflect a common understanding of what Web 2.0 means, the question of how such a concept came to life and evolved since its first use remains unanswered.

In this research we address this question by examining the history of the Wikipedia entry for Web 2.0. More precisely, we aim to answer the following questions: How did the Wikipedia community construct the Web 2.0 concept over time? Who are the members of this community? What are their roles and behaviors? How did the content of the Web 2.0 page evolve and why? Our motivation to choose Wikipedia as the object of study was twofold. First Wikipedia is an instance of Web 2.0 applications, and since its creation in 2001, Wikipedia has grown into one of the largest reference Web sites.

Beyond its attempt to reveal how Wikipedia organized the vision for Web 2.0, this micro-level study should also contribute to the research that seeks to understand the mechanics of the Wikipedia community participation and content building [4], [5].

## 2. METHODOLOGY

To answer our questions, we looked at the evolution of the Web 2.0 entry in Wikipedia over three years, from the first entry in March 2005 until February 2008, importing the page edits history for these three years from Wikipedia. The revision history dump was downloaded as an XML file and then imported into Excel for analysis.

The revision history provided key information about each edit: when it was made (timestamp), who made it (username or IP address), a short summary describing the modification performed if any, and the content of the page format in textual form. Analysis of this information was simple but tedious, and consisted

of examining each edit individually to classify them according to specific types, to identify the various type of contributors (administrators, bots, registered or unregistered users), calculating the size (number of characters) of each entry, and other useful metrics, such as the number of characters added or subtracted, the time interval between two consecutive edits, etc. In what follows we report the results of the quantitative analyses we carried out.

### 3. ARTICLE CONTENT EVOLUTION

Over the 1,071 days studied, the page was edited 3,665 times (an average of 3.4 edits a day), and its content grew from a small entry containing 640 characters to a fully developed page containing over 28,000 characters (see Figure 1). The size of the Web 2.0 article grew steadily over time, not without dramatic episodes as revealed by the number of isolated data points on the graph, and seems to have reached a plateau after the article's first 3 years of existence.

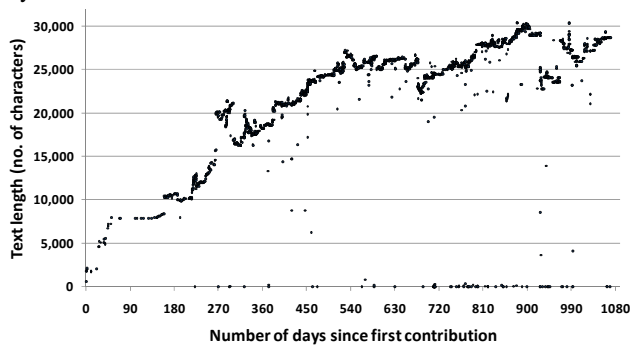


Figure 1. Web 2.0 article growth

Our categorization of the types of edits that made up the content of the article reflects the contributors' purposes. Some of the categories we used already existed in the Wikipedia glossary (vandalism, spam, etc.), while others were created for the purpose of this study. They are shown in Table 1.

Table 1. Types of edits performed to the Web 2.0 article

Edit types	Number of edits
Unchallenged	1,206 (32.9%)
Challenged	701 (19.1%)
Challenge	541 (14.8%)
Restoration	285 (7.8%)
Copy edit	267 (7.1%)
Vandalism	232 (6.3%)
Test	185 (5%)
Maintenance	176 (4.8%)
Spam	72 (2%)
Grand Total	3,665 (100%)

Together vandalism and spam represented 8.3% of all edits to the page. Clean-up of either type of edits was categorized as "restoration" edits. Some of these cleaned up various vandalism or spam attempts, explaining the discrepancy between the number of vandalism and spam edits and the number of restoration edits. All non-spam and non-vandalism edits that were "reverted" or

"undone" were categorized as "challenged" edits, to emphasize the perceived controversy surrounding such edit.

Edits that reverted or undid previous edits were categorized as "challenge" edits. Some "challenge" edits reverted several "challenged" ones, explaining the difference between the number of "challenged" and "challenge" edits. "Challenge" and "Challenged" edits represent 33.9% of all edits performed to the article, and reflect in some ways the tensions among the various opinions that prevailed about what the Web 2.0 concept should have been.

"Unchallenged" edits refer to those edits that were not challenged by either a revert or an undo action. These edits are important because they were not dismissed immediately, and as such formed a more stable fabric upon which further content could be built. "Unchallenged" edits represent the bulk of the Web 2.0 article contributors' activities (32.9% of all edits).

"Maintenance" edits are edits that insure the validity and integrity of external and internal links, perform disambiguation, etc., while "copy edits" are minor edits that fix grammatical typographical errors. These edits essentially reflect the attention paid to the quality of the article, according to Wikipedia standards. At times, some contributors edited the page and immediately deleted their contribution in a consecutive entry. These edits were categorized as "test" edits. All together, "test", "maintenance" and "copy" edits represent 16.9% of all edits.

As the article evolved, so did the combinations of types of edits performed, as shown in Figure 2. The first year of the article's existence saw little controversy about its content, as the small number of "challenge" and "challenged" edits reveals. During this period most edits were of the "unchallenged" type, growing the article to 70% of its final size.

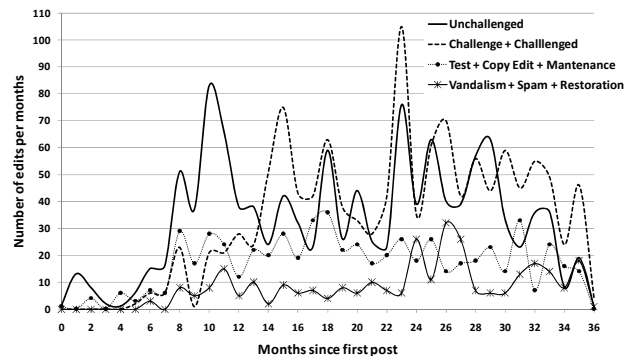


Figure 2. Evolution of types of edits with time

Real attention to the quality of the article, as expressed by the number of "maintenance" edits and copy edits, started 6 months after the first entry, probably triggered by a burst of "unchallenged" edits during this period, and increasing visibility of the article. Vandalism, spam and their restoration peaked 23 months after the first contribution, suggesting that the article reached substantial visibility at this stage.

In summary, the Web 2.0 article grew in a complex manner. Leaving aside the inevitable vandalism and spam, the evolution of the combinations of different types of edits revealed interesting patterns which we will explore momentarily. Our attention first turns to the individuals who contributed to the article, and we explore next who they were and what they did.

## 4. CONTRIBUTORS TO THE ARTICLE

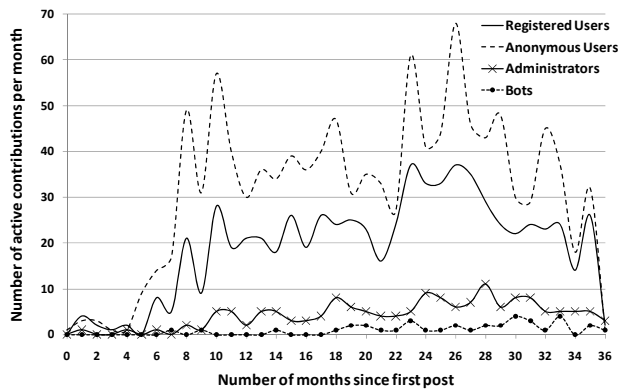
Over our period of study, 1,777 users contributed to the article (see Table 2 for details). Anonymous users, those unregistered users only identifiable by their IP addresses, represented the largest group (1,126 in total). Registered users, users who set an account with Wikipedia, formed the second largest group (540 in total), followed by administrators, elected registered users with special editing rights (99), and bots which perform automated tasks (12).

**Table 2. Types of users contributing to the article**

User types	Number of contributors	Number of edits
<b>Anonymous users</b>	1,126 (63.4%)	1,700 (46.4%)
<b>Registered users</b>	540 (30.4%)	1,603 (43.7%)
<b>Administrators</b>	99 (5.6%)	294 (8%)
<b>Bots</b>	12 (0.7%)	68 (1.9%)
<b>Total</b>	<b>1,777 (100%)</b>	<b>3,665 (100%)</b>

In terms of number of edits, anonymous and registered users contributed the most and almost equally to the Web 2.0 article. Administrators and bots performed very small number of edits in comparison.

Anonymous and registered users were both active from the article inception, and the ratio of anonymous to registered users oscillated around 2 over the period of study. Administrators and bots became active later, 9 and 18 months after the first contribution, respectively (see Figure 3).



**Figure 3. Types of contributors active over time**

Different types of users performed different types of edits. Not surprisingly, anonymous users were the biggest contributors of vandalism (93.5%) and spam (80%), which were restored in most part by registered users, and to some extent by administrators and bots. Anonymous and registered users were the biggest contributors of “unchallenged” edits, contributing almost equally (49% and 47.6% respectively). Anonymous users’ contributions were the most “challenged,” (73% compare to 25% for registered users). Those edits were challenged mostly by registered users (61.6%) and administrators (28.1%).

## 5. FOUR PHASES OF EVOLUTION

The patterns of evolution of the types and numbers of contributors and edits lead us to propose four major periods in the evolution of

the Web 2.0 article: seeding, germination, growth and maturity (see Figure 4).

### 5.1 Seeding

The first entry was made by an anonymous user who planted the seed of what would become a fully fledged article. During this phase (the first 6 months of the article’s existence), the page developed slowly (the average time between consecutive edits was 83 hours), being edited 12 times per month on average. Few contributors were involved: 7 users were active each month on average during this period. There was no activity from bots or administrators during this phase. Anonymous users were twice as many as registered users, and both contributed mostly “unchallenged,” edits to the content of the article. There was no vandalism or spam, which suggests that the article still received little attention at this stage, the seed being underground. Each month during this period, an average of 1,871 characters were added and 342 characters were subtracted every month, making the article 6,379 characters long at the end of its first 6 months of life.

### 5.2 Germination

During the Germination phase (7 to 21 months after the first contribution), the article’s growth took off. Editing intensified during this period with an average of 110 edits per month from 12 during the previous period. The average time between consecutive edits decreased from 83 hours to 8. The page attracted more contributors, and the average number of active users per month rose from 12 to 62. Anonymous users were still twice as many as registered users, and a handful of administrators joined the crowd. As the page received more attention, it began to be subject to vandalism and spam, all performed by anonymous users and restored by registered users and administrators. Anonymous users were much more than just vandals during this period, however. While their contributions began to be challenged by registered users and administrators, they carried on, providing the majority of “unchallenged” edits (57.5% of the total). Maintenance and copy editing activities also intensified greatly during this period, suggesting that increasing attention is paid to the quality of the article. Each month during this period, an average of 34,174 characters were added and 33,139 characters were subtracted, making the article over 20,000 characters long at the end of this period.

### 5.3 Growth

The next phase, growth (22 to 29 months), was the most active period of development. The page was edited 156 times a month on average, and the average time between two consecutive edits decreased to 5 hours. The page attracted an increasing number of contributors, with 87 users active on average each month. Unregistered users remained the biggest group of contributors, while registered users increased their presence further. More administrators and bots were also active. Vandalism and spam activities increased twofold from the previous phase, and originated mostly from anonymous users. Weeding out vandalism and spam became a major activity of registered users, bots and administrators. Registered users became the biggest contributors of “unchallenged” edits (57% of the total). Together with administrators, they also increasingly challenged anonymous users’ contributions.

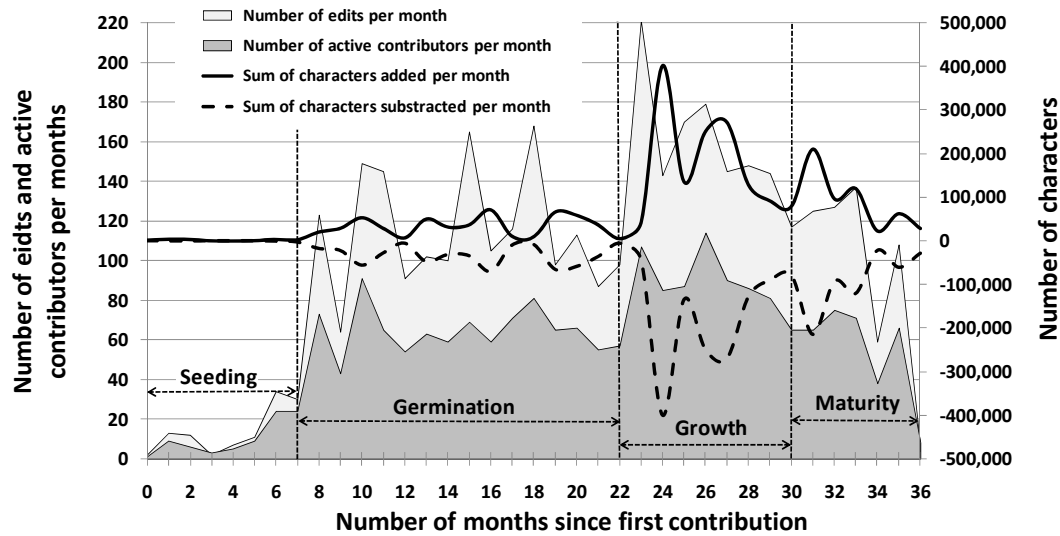


Figure 4. The four phases of evolution

Each month during this period, an average of 165,861 characters were added and 165,449 characters were subtracted, making the article over 25,000 characters long at the end of this period.

## 5.4 Maturity

During the last phase (30-36 months) the article appeared to reach maturity. This period saw a rapid decrease in contributors' presence and editing activities. The article was edited 97 times on average each month, down from 156 the previous period, and the time between consecutive edits rose from 5 to 8 hours. Fewer contributors were active, an average of 57 each month, mostly pruning a now fully developed article. The number of "unchallenged" edits being performed dropped by 60%, and vandalism and spam decreased by 25% from the previous period, bots taking the lead in dealing with these. This lack of attention suggests a loss of interest from actual and potential contributors to the article.

The monthly average number of "challenge" and "challenged" edits decreased by 37%, to a level close to that of the Germination phase, suggesting that the loss of attention to the article may be due to the fact that a consensus about what the concept of Web 2.0 means has been reached. Each month during this most recent period, an average of 88,499 characters were added and 88,624 characters were subtracted, making the article over 28,000 characters long at the period's end.

## 6. CONCLUSION

In this paper, we investigated the evolution of the Wikipedia entry for Web 2.0 over the three years since its inception in March 2005, with the aim of furthering our understanding of how such a concept develops over time. We categorized the patterns of evolution of the types and numbers of contributors and edits into 4 distinct phases: seeding, germination, growth and maturity. This simple stage model provides new insight into how a Wikipedia article about a buzzword such as Web 2.0 develops over time. How such a growth model applies to other types of Wikipedia articles should provide interesting avenues for future research.

Many more pressing questions remain to be answered, however. Although we have established who did what and when, these relationships were only expressed in terms of number of edits, which begs the question of *what* was actually contributed and when. Who the biggest contributors were, in terms of the size of their contributions to the page, should also be investigated further. In all, answers to these questions should provide a more detailed picture of the community who participated in the building of the Web 2.0 article.

## 7. ACKNOWLEDGMENTS

This project was funded in part by the UCLA Anderson School's Information Systems Research Program.

## 8. REFERENCES

- [1] Swanson, E. B. and Ramiller, N. C. 1997. The organizing vision in information systems innovation. *Organization Science* 8 (5):458-474.
- [2] Wang, P. and Swanson, E. B. 2007. Launching professional services automation: Institutional entrepreneurship for information technology innovations. *Information and Organization* 17:59-88.
- [3] Wang, P. and Swanson, E. B. 2008. Customer relationship management as advertised: Exploiting and Sustaining Technological Momentum. *Information Technology and People* 21 (4):323-349.
- [4] Viégas, F. B., Wattenberg, M., and Dave, K. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vienna, Austria April 24-29, 2004)*. CHI '04. ACM Press, New York, NY, 575-582.
- [5] Kittur, A., Chi E., Pendleton, B. A., Suh, B., Mytkowicz, T. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *Alt.CHI at CHI 2007, 2007 April 28-May 3, San Jose, CA*.