A Dive Into Online Community Properties

Patrick Wagstrom
IBM Research
19 Skyline Dr
Hawthorne, NY 10532 USA
pwagstro@us.ibm.com

Jacquelyn Martino
IBM Research
19 Skyline Dr
Hawthorne, NY 10532 USA
jmartino@us.ibm.com

Juerg von Kaenel
IBM Research
19 Skyline Dr
Hawthorne, NY 10532 USA
jvk@us.ibm.com

Marshini Chetty*
College of Computing
Georgia Institute of Technology
85 5th Street NW
Atlanta, GA, 30332 USA
marshini@cc.gatech.edu

John Thomas
IBM Research
19 Skyline Dr
Hawthorne, NY 10532 USA
jcthomas@us.ibm.com

Lauretta Jones*
12 Carolyn Way
Purdys, NY, 10578 USA
laurettajones@mac.com

* - Work performed while at IBM Research

CSCW 2011, March 19–23, 2011, Hangzhou, China.
ACM 978-1-4503-0556-3/11/03.

Abstract
As digital communities grow in size their feature sets also grow with them. Different users have different experiences with the same tools and communities. Enterprises and other organizations seeking to leverage these communities need a straightforward way to analyze and compare a variety of salient attributes of these communities. We describe a taxonomy and tool for crowd-sourcing user based evaluations of enterprise relevant attributes of digital communities and present the results of a small scale study on its usefulness and stability across multiple raters.

Keywords
Online Communities, Taxonomy, Enterprise, Visualization

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. H3.5. Online Information Services: Web-based Services.

General Terms
Human Factors, Management

Introduction
Digital communities enable user-to-user collaboration and self-organization on an amazing scale. While open projects, such as Open Source software, can utilize these communities and tools to collaborate and coordinate in a very public way, enterprise firms often have information security requirements that preclude

adoption of such public tools. As a direct consequence enterprise software firms often have to re-implement digital communities for use behind a secure corporate firewall. For example, many corporations utilize private instant messaging systems[4] and, increasingly, private social networking and communication systems that mirror common communities such as Facebook and Twitter[2].

Failure to understand properly all of the aspects of a community and the requirements it imposes on software development can result in an unsatisfactory experience and project failure. For example, a single user may not utilize a critical feature of the community and therefore not include it in the design of an enterprise version. Some features may require a larger or different style of community than the private enterprise can adequately support; such as the bots that continual massage Wikipedia. The enterprise also may be ill-suited to design and deliver a private version of the tool because of larger issues, such as a lack of expertise in a critical area or regulatory issues that prevent implementation of critical features, for example, anonymity.

To mitigate the risk in these actions, we developed a taxonomy for the evaluation of online communities within the context of enterprise software development. We then describe a novel method for users to evaluate communities against the taxonomy. Finally, we present a small study of the tool and taxonomy to evaluate the stability and usability of the taxonomy.

A Taxonomy of Digital Communities

As part of a long-term research project we developed a taxonomy of online community attributes applicable to a wide variety of different digital communities. Rather than seeking to be a general taxonomy as proposed by Lazar and Preece[3] or Dubé et. al.[1], we focused on attributes most relevant to enterprise development and use. It is organized into a six different axes that address primary aspects online communities:

- **Purpose:** the set of reasons why people participate in the community.
- **Economy:** the types of goods, services, artifacts, and reputation related elements are collected, shared, and traded in the community.
- **Member:** the people that the site targets and how much information the community collects about community members.
- **Platform:** the technical features of the platform on which the community is constructed.
- **Content:** issues around the content of the community such as the origin and ownership of the content and the ability of community members to interact with and modify the content.
- **Interaction:** ways that community members can express their intentions, directly collaborate, and other features related to individual and group interaction patterns.

Each of the axes is further divided into a number of different voices that address various sub-issues. For example, the Platform axis has voices for the presence of an API, security features of the community, tools to access the community, and the timeliness for messages sent to community members.

Each voice may be one of three different types: multiple choice, scalar, or percentage. Multiple choice

is used when the responses have no logical progression – such as an enumeration of tools that can be used to access the community. Scalars are used for voices with a logical progression – such as the timeliness of messages. Percentage voices are unique to the Purpose axes where they are used to sum to 100% across a variety of different voices.

Each side of the visualization represents an axis of the taxonomy. Each voice is a slice within the axis. Smaller partitions distinguish between different values. The visualization is filled in as the user enters data and rotates to present to indicate which axis and voice the user is currently addressing.

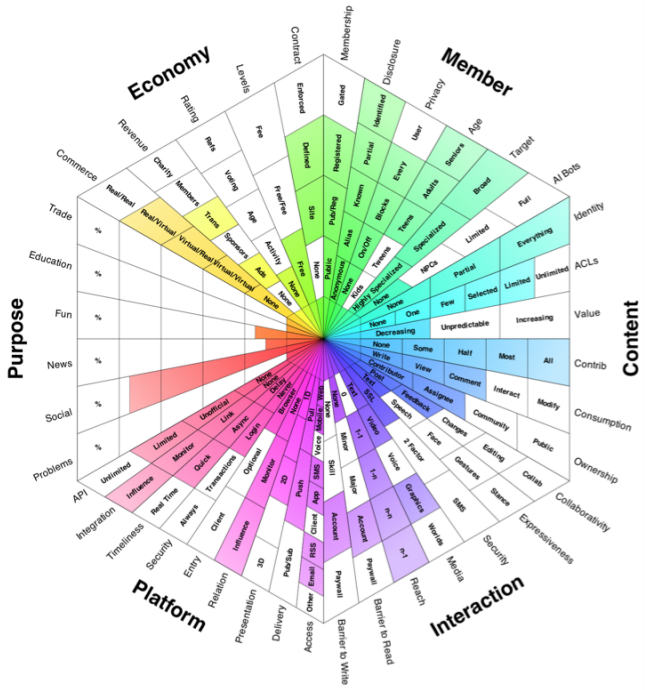

figure 1: The taxonomy visualization is filled in and rotates as the user completes an evaluation. This makes it easy to compare results within and across communities.

Within our taxonomy there are six axes, 34 different voices, and 100 different possible voice values (excluding percentage voices). In a traditional system this would be presented as a set of questions in a linear fashion with little connection between related elements. To address the possible monotony and lack of context from such a system, we developed an innovative visualization and method for presenting the questions, as seen in figure 1. This visualization and input tool, dubbed CommunityScout, also makes it easy to quickly compare results between and across communities.

Taxonomy Study and Validation

To evaluate the tool and community taxonomy we conducted a study with students and researchers familiar with a wide variety of online communities. Participants were given a choice of community to evaluate and each participant could evaluate more than one community.

A total of 26 individuals were invited to participate in evaluating CommunityScout. 14 individuals visited CommunityScout and submitted 19 community evaluations. Facebook was the most popular community evaluated with evaluations from four users. Other communities with multiple evaluations were LinkedIn, Reddit, and World of Warcraft.

For each of the communities with multiple evaluations the two-way intra-class correlation for consistency was calculated. It ranged from a high of 0.964 for Facebook to a low of 0.757 for Wikipedia. In all cases the result was highly significant, largely due to the number of data points evaluated for each community.

table 1: Intra-class correlation coefficients for communities with multiple evaluations

Community	ICC	95% Interval
Facebook	0.964	0.950-0.975
LinkedIn	0.826	0.767-0.875
Reddit	0.836	0.757-0.891
Wikipedia	0.757	0.646-0.836
World of Warcraft	0.943	0.920-0.961

In addition to calculating the ICC across the taxonomy, we calculated the ICC for each of the axes within each community. We found that the content axis consistently had the lowest ICC of all the axes. While it was still significant in all cases there were several cases when the confidence interval included the null hypothesis, such as Facebook (p=0.04). Examination of this axis found that there was disagreement in the questions about the ownership of content, methods of interacting with user created content, and whether or not content became more or less valuable over time.

table 2: Intra-class correlation coefficients for axes in evaluations of Facebook

Axis	ICC	95% Interval
Content	0.336	-0.040-0.822
Economy	0.854	0.704-0.946
Interaction	0.824	0.702-0.913
Member	0.814	0.601-0.942
Platform	0.872	0.685-0.955
Purpose	0.898	0.696-0.983

The reason for the inconsistency around the content axis could be for one of two reasons – either there is a lack of clarity in just this axis, which is possible but unlikely given the high ICC of the other axes, or many community participants are unaware of all of the features of a community. In particular, this seemed to be apparent surrounding the issue of who actually owns and has rights to content submitted by users.

Further qualitative responses from study participants indicated that they believed the taxonomy and visualization were helpful in understanding the attributes of a community, particularly as it made apparent many attributes of a community that are not immediately obvious to the casual observer (mentioned by two different participants). We hope to further expand the tool to evaluate more communities and better understand how these non-apparent attributes affect the ability of an enterprise to learn from existing online communities.

Citations

[1] Dubé, L., Bourhis, A., and Jacob, R. Towards a typology of virtual communities of practice. *Interdisciplinary Journal of Information, Knowledge, and Management*, 1:3-13, 2006.

[2] Ehrlich, K. and Shami, N.S. Microblogging inside and outside the workplace. In *Proc 2010 AAAI Conference on Weblogs and Social Media*. AAAI, 2010.

[3] Lazar, J. and Preece, J. Classification schema for online communities. In *Proc AMCIS 1998*

[4] Quan-Haase, A., Cothrel, J., Wellman, B. Instant messaging for collaboration: a case study of a high-tech firm. *Journal of Computer-Mediated Communication*, 10(4)2005.