Enabling Social Translucence for Wikipedia

Travis Kriplean1, Mark Zachry2, Alan Borning1, David W. McDonald3

1Computer Science & Engineering
2Human Centered Design & Engineering
3The Information School
University of Washington, Seattle, WA 98195
{kriplean,zachry,borning,dwmc}@uw.edu

ABSTRACT
Newcomers to online communities often have trouble understanding the context and activities of other participants. But this is not just a problem to newcomers; even established members can struggle to comprehend social dynamics in their own communities. Social translucence is a conceptual approach to designing systems that facilitate better sensemaking about others and their actions in online social settings. We have been developing a framework to facilitate socially translucent tools and extensions for distributed contributor communities such as Wikipedia. This demo presents example tools and an evolving architecture to illustrate characteristics of social translucence.

Author Keywords
Wikipedia, social translucence, online community.

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Computer-supported Cooperative Work.

General Terms
Design, Human Factors.

INTRODUCTION
Trent muses to himself that he always seems to get the most difficult cases. He was recently asked to mediate a dispute between people editing the Anarchy page on Wikipedia. Not only is the page very active with article edits and discussions about those edits, but many of the participants are also individuals with widely varying, strongly held opinions. When reviewing the article’s history, Trent can see a series of edits, and numerous page reverts, but getting a clear sense of the activities of the editors is hard. If only there were a way to get a sense of who was involved, how the various factions were interacting, and to tease out the content disputes from the various behavioral issues [1][2].

Understanding the behaviors of others in an active online community can be difficult, for newcomers and experienced members alike. Socially translucent systems provide mechanisms to help see and understand activity not only as it is taking place, but activities in the past that may influence people’s current actions. Indeed, in situations like Trent’s, understanding which actions may have led to a disagreement is a critical aspect of the context necessary for mediating workable resolutions.

Our research focuses on tools and infrastructure to facilitate social translucence in online communities. We take Wikipedia, with its rich logs of user activity, complex division of work, and often divergent user opinions, as one case that could benefit from increased translucence. Though our focus here is on Wikipedia, we are also employing similar concepts in the design of citizen-government engagement systems.

SYSTEM ARCHITECTURE
An architecture for social translucence requires components that collect information, expressions that support composition of activities, and interfaces that visualize those activities (Figure 1). Our framework has five levels. The lower levels facilitate finding, mining, and exposing salient social activities. The upper levels of the framework will allow community members to compose and integrate representations of salient activities into useful tools. This CSCW demonstration focuses on tools at the upper levels.

At the base level, datasets provide an information source. This includes the full edit history on Wikipedia, as well as page view data. Although such data is potentially available to all system users, such raw data is very difficult for many users to understand or interpret. A feature extractor parses the...
data to extract potentially valuable tokens. One or more of these tokens comprise primitive features such as reverses, administrative actions such as page locking, co-editing activities, barnstars, text persistence and policy citations. Such features can be treated as indicators of higher-level, socially salient activity.

A semantic interpreter transforms primitive features into socially salient compound features, such as roles based on a model of work, social network structures, and maps of conflicts amongst users and pages. The semantic interpreter links the observable, extracted activities into social constructs. Such linking requires an understanding of how Wikipedians understand the observable actions of other Wikipedians.

All of the architectural components described so far help to find and represent social activity. The rest of the architecture is oriented toward exposing and supporting user generation of tools that enable greater social translucence.

The reflexive composer exposes representations of the compound and primitive features through a reflexive composition language. Community members and developers use the language to express (compose) their own meaningful combinations of relevant features. The language will need to be more complex than a markup language but somewhat less complex than a traditional scripting language. The reflexive composer allows the specification of application modules.

Finally, the component visualizer links one or more application modules to a meaningful user presentation. The visualizer requires (1) a persistent representational model of the application modules as defined by the reflexive composition language and (2) a means to integrate these representations into working tools. Our efforts here are to develop a basic declarative syntax. The ultimate goal, however, is to build, and enable the community to build, applications that support social translucence.

SENSEMAKING INTERACTIONS

The current system facilitates interactions at two different levels. We have developed a rudimentary reflexive composer that takes simple declarations and applies them to a specific page context through Wikipedia’s template mechanism. Further we have developed a component visualizer to make them conveniently accessible. Though these initial applications are straightforward, they provide grounding for discussions about fostering translucence for online interactions.

Reflexive Composer

Users can currently construct templates in their personal user spaces that define various activities that should be displayed, and how to visualize them (Figure 2). These templates can also be defined in communal spaces so that a group can come to agree on important activity dimensions. Like any template, editing the template can result in different features being passed to a page when a page is rendered in the browser, as the user traverses the encyclopedia. These features are passed to the visualization component. The system leverages MediaWiki’s powerful templating system in order to parameterize definitions and support inheritance. Templates also provide an easy mechanism for gathering all of the defined templates of a given type. The proper template can then easily be loaded, based on the user and context, regardless of where the template is defined.

Component Visualizer

The component visualizer currently shows basic statistics invoked by mouse hovers. Other supported visualizations on mouse hover include TF-IDF weighed word clouds that show, e.g., the individual words that comprise the barnstars a user has received, or the words that comprise a users’ edit summaries. The visualizer currently examines the appropriate user page for the set of defined templates to determine which visualization and which features to display. Figure 3 is an example, using the template in Figure 2.

The system currently runs through a reverse proxy server of Wikipedia, which injects javascript into the requested pages. The Javascript uses ajax calls to pull data from a Django-based API whenever a relevant context is entered. The API itself pulls data from a MySQL database containing datasets derived from a full parse of the February 2008 English Wikipedia dump. On page load, the Javascript also identifies and loads all relevant templates.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation award number IIS 0811210.

REFERENCES
