Wiki Architectures as Social Translucence Enablers

Stephanie Gokhman
Human Centered Design & Engineering
University of Washington
Seattle, WA
sgokhman@uw.edu

David W. McDonald
Information School
University of Washington
Seattle, WA
dwmc@uw.edu

Mark Zachry
Human Centered Design & Engineering
University of Washington
Seattle, WA
zachry@uw.edu

ABSTRACT
Whether novice or expert, it is useful for contributors to understand the environment to which they are contributing, including the relationships of other users to the content and to users. However, the relationships and work that enable content creation in an online contributor system, such as Wikipedia, are not always visible. To expose and better understand these relationships, we have built an information visualization toolkit called Re:Flex to support components of social translucence in Wikipedia, with broad applicability to other contributor systems. By mimicking the flexible, fluid architecture of a wiki within the blackboard architecture of this visualization toolkit, we demonstrate how the composable interactions inherent to contributor systems can be mirrored in the tools that support the work which creates them.

Categories and Subject Descriptors
H.5.3 [**Group and Organization Interfaces**]: Computer Supported Cooperative Work

General Terms
Design, Theory.

Keywords
Wikipedia, software architecture, social translucence, wikis

1. INTRODUCTION
Online contributor systems like Wikipedia contain significant "hidden work," where relationships between contributors and content are discrete. By exposing these hidden relationships, contributors can distinguish and build from knowledge of the activity patterns of other users, aid in decision-making, and facilitate cooperative processes [5], such as advice-seeking, consensus-building, determining trustworthiness of content or expertise of a contributor. Exposing these relationships follows the theoretical perspective of social translucence, which has three critical components: (a) mutual awareness of activities in the community, (b) propagation of contextually salient cues, and (c) accountability for actions taken [2].

Like many socio-technical constructs, social translucence is not something that can simply be pasted or bolted on. Designing for social translucence begins at the level of software architecture. The underlying software architecture of a system facilitates the observation of different types of activities within the system, the communications among those processes, and the data that is stored and retrieved—all at the behest of some user action. Our primary contribution draws parallels between the blackboard (or blackboard-like) architecture of most online contributor systems and the architecture of tools to support the collaborations of contributors through social translucence. Further, we illustrate a specific software architecture that we are building that could complement many different types of wikis and that could simplify the construction of a range of tools and visualizations that support social translucence.

WikiSym'11, October 3–5, 2011, Mountain View, California, USA.
ACM 978-1-4503-0909-7/11/10.

2. ARCHITECTURE FOR TRANSLUCENCE
Contributor systems, and in particular wiki systems, are built with blackboard or blackboard-like architectures. The software architecture community considers blackboards to be a generic architectural style sometimes called a "repository" or "data-centered" style. This architectural style facilitates complex state sharing, while loosening the constraints of the formal blackboard system [6]. While initially "blackboard" defined a strict rule-based structure [4], new adaptations allow for a flow of control that determines the state of the data on the blackboard. Our visualization toolset is built on this style of data-centered architecture, wherein various processes alter the state of a blackboard: extracting data, designating patterns within the data, and posting this data in a new form. Mimicking the flexible control flow of a wiki, user interactions with our toolset define the state of the blackboard to enable social translucence. An architectural view of our extension is in Figure 1. The following components in aggregate describe the architecture of our social translucence visualization toolkit.

2.1 Information Source
The data within the blackboard is an aggregate itself of a number of Wikipedia data sources. These foundational data sets are the conductors that transmit information into the blackboard.
Wikipedia Dumps: These are the regular dumps of Wikipedia data, cleaned and processed in order to populate our system's data objects and abstractions.
Wikipedia Special Pages: These pages feed into the social proxy and provide the content base under which our visualizations lie.

2.2 Blackboard
The blackboard is the main construct of our system, which controls and contains all global data, performs calculation and problem-solving before build time, inspects data, performs abstraction and updates the visualizations. This structure includes a number of components that together provide content to the system.

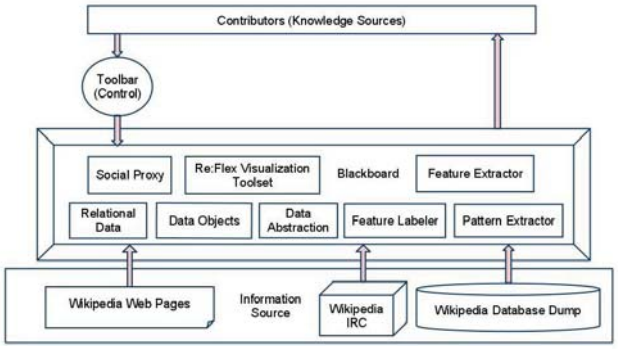

Figure 1. Architecture of Re:Flex uses the information source to populate the data of the blackboard, which displays information based on the contributors' selection in the toolbar.

Wikipedia IRC Streams: Wikipedia supports an Internet Relay Chat (IRC) channel where updates to the system are announced within the chat text. A process listens to one or more channels and posts updates to the blackboard. This stream provides real-time content to update the original database dump.

Social Proxy: This is the set of webpages proxied from Wikipedia.

Visualization Toolset: This is the set of visualizations that lie on top of the social proxy.

Relational Data: This is the primitive set of relations among data objects, drawing connections between users and particular pages, largely through aggregated edit counts within set time frames.

Data Abstraction: This is the set of high-level calculated relations among data objects and relational objects in the database that demonstrate activity such as interacting on user talk pages, vandal-fighting, and traditional editing contributions.

Feature Extractor: Similar to the social proxy, the feature extractor reads Wikipedia pages; however, this piece also processes and extracts features that are not properly part of the Wikipedia database dump. These raw features provide basis for collection of headings, section contributions, template tags, and barnstars.

Feature Labeler: Building from the feature extractor, the feature labeler appends to the raw features and annotates them in order to develop meaning. This label is generated through a higher-level processing depending on the types of features identified through machine learning, crowdsourcing or human computation. For example, the feature extractor may provide a number of barnstars, but the feature labeler will determine whether these barnstars were awarded for negative or positive contributions.

Pattern Extractor: Building from the feature extractor and labeler, this extractor looks for patterns or chains of activity for an individual user or for groups of users, such as WikiProject participants. This process identifies and collects instances of possible patterns from the set of features and posts them back to the blackboard. This piece is of particular interest as it holds both value to the users and to our research, demonstrating patterns of participation in open collaboration.

2.3 Contributors as Knowledge Sources

Specialists, in this case Wikipedia contributors, operate the blackboard by specifying conditions, performing manipulations and executing actions. These are the users who operate the controls that then process and visualize the data set. The knowledge sources, in this way, are the catalyst for the operation of the visualization toolkit.

2.4 Toolbar Control

The toolbar of our visualization toolset provides the options that the knowledge sources can select to define the data source, loop the data source, and end the session.

3. CONCLUSION

Through integration of a number of key components of traditional blackboard architectures, our system offers tools to support social translucence in complex collaborative efforts oriented toward creating content in systems like Wikipedia. The research work detailed within this poster provides specific insight into the significance of social translucence in online contributor systems as well as a technical means for supporting social translucence by mimicking interactions inherent to these environments.

4. ACKNOWLEDGMENTS

This work is supported in part by National Science Foundation Grant NSF IIS-0811210.

5. REFERENCES

[1] Corkill, D. D. (1991). Blackboard System. *AI Expert*, 6(9), 40-47. doi: 10.1016/0950-7051(89)90039-7.

[2] Erickson, T. and Kellogg, W.A. Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Transactions on Computer-Human Interaction*, 7 (1). 59-83.

[3] Erickson, T., Halverson, C., Kellogg, W., Laff, M. and Wolf, T. Social Translucence: Designing Social Infrastructures that Make Collective Activity Visible. *Communications of the ACM*, 45 (4). 40-44

[4] Hayes-Roth, B. A blackboard architecture for control. *Artificial Intelligence*, 1985, 26, 251-321.

[5] Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information Visualization*, *7*(1), 492. doi: 10.1057/palgrave.ivs.9500167.

[6] Shaw, M. and Garlan, D. *Software Architecture: Perspectives on an Emerging Discipline*, 1996. Prentice-Hall, Inc.