genSpace: Exploring Social Networking Metaphors for Knowledge Sharing and Scientific Collaborative Work

Chris Murphy, Swapneel Sheth, Gail Kaiser, Lauren Wilcox
Columbia University

Introduction

- Scientists collaborating together in the same lab on the same project share:
  - Data: specimens, samples, materials, observations, etc.
  - Tools: instruments
  - Knowledge: open discussion
- However, there are time and space constraints
- More significantly, this does not scale well to **communities** of scientists working on different projects but who could possibly learn from each other's expertise, experience, etc.

CSCW Approaches

- CSCW aims to augment same-time/same-place collaboration but more significantly different-time/different-place collaborations and communities
- Current generation CSCW systems support *data sharing* (e.g. Biological Systems Collaboratory) and/or *tool sharing* (e.g. BioCoRE)
- However, these systems do not address **knowledge sharing**
  - how/when/where/why to use tools and data

Knowledge Sharing

- Knowledge sharing is partially enabled through static approaches: publications, email lists, message boards, wikis, etc.
- We seek to add **knowledge sharing** to CSCW, but without requiring "extra work" on the part of scientists

Social Networking

- Some online social networking is a form of CSCW that is potentially enjoyable and profitable but still requires “extra work”
  - Facebook, MySpace, LinkedIn, etc.
- Other social networking implicitly records what people do online to aggregate, data mine, disseminate in an enjoyable and profitable fashion, but with no “extra work”
  - Collaborative filtering

Overview

- We combine implicit and explicit social networking concepts in our approach to augmenting data and tool sharing CSCW with knowledge sharing
- We present a prototype implementation of such a system in the domain of computational biology
- We examine some of the software engineering implications of such an approach

Background

- We are working with Columbia University's Center for Computational Biology & Bioinformatics to enable collaboration and communities
- Many computational biologists use geWorkbench, a platform for analysis and visualization tools for integrated genomics
- geWorkbench is standalone and has no collaboration facilities

Important Questions

- What analysis tools should I use to investigate this problem?
- Who do I know who also uses this tool?
- Which tools work well together?
- Where does this tool fit in a typical workflow?
- When did I previously use this tool?
- How can I get help (from an expert who is online right now)?

Analogous Social Networking Questions

- What movies would I like?
- Who also likes this book?
- Which food and wine go together?
- Where does this song fit in a playlist?
- When was this restaurant last reviewed?
- How can I get help about this MP3 player?

Approach

- The use of tools in the domain indirectly encapsulates knowledge and expertise
- We automatically build organizational and community memory by monitoring what users do with a particular set of tools
- Aggregated logs can then be mined and knowledge can then be exposed via social networking models

Implementation

- **genSpace** is a set of plugin components for geWorkbench
- Instrument geWorkbench to capture and record analysis events
- Aggregate event logs for communities of users
- Data mine event patterns and then expose them via social networking functionality

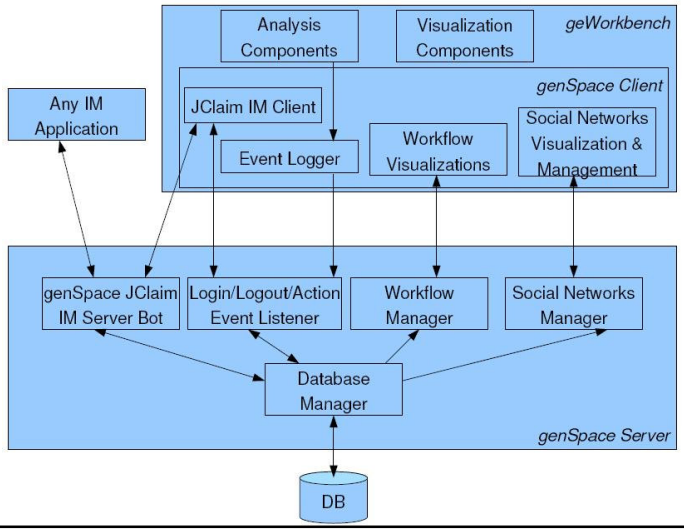
genSpace Features (1)

- **Social Networking**: allows users to associate with each other and share knowledge
- **Collaborative Workflow Composition**: past history of analysis tool usage is used to identify and visualize commonly-occurring sequences/workflows
- **Instant Messaging (IM)-based User Interface**: communication with the system may be done via IM clients such as Yahoo! Messenger, Windows Live Messenger, or Google Talk

genSpace Features (2)

- **Peer Suggestions**: suggests other genSpace users who work with similar analysis tools
- **Tool Suggestions**: suggests analysis tools that may be useful, based on what tools were previously used
- **Expert Finder**: identifies genSpace members who appear to be experts in using geWorkbench, a particular analysis tool, or a set of tools

genSpace Architecture



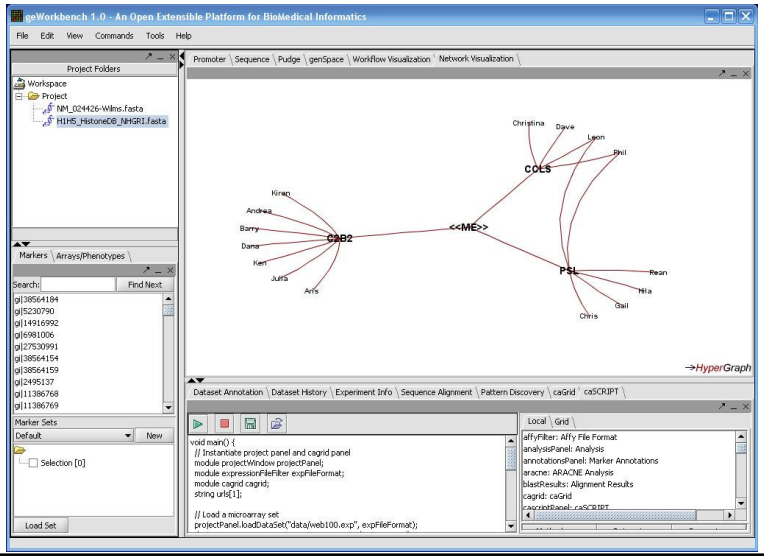
Privacy Concerns

- Users can choose anonymous logging or disable it entirely
- Users can also opt out of being suggested as expert or peer
- Security of the activity logs will also need to be investigated

Social Networking

- Users can list, create and join social networks
- Also create lists of friends and see who is currently logged in
- Can also query for the most popular tool, either in social network or all genSpace

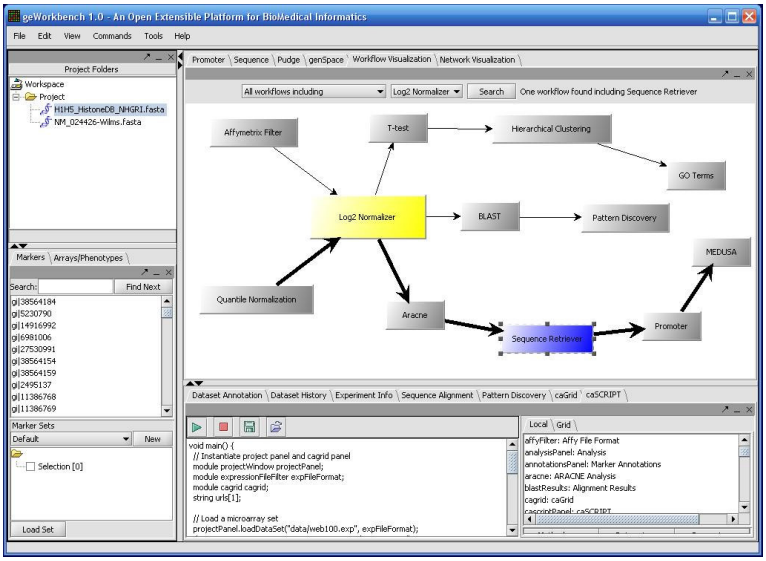
Social Network Visualization



Collaborative Workflow Composition

- Addresses the challenge of selecting from a number of analysis and visualization tools
- System investigates what other workflows have been performed in the past by other users and then suggests one or more “workflows” based on other users’ previous activity

Workflow Visualization



Peer and Tool Suggestions

- "Friend finder" allows users to find others who have similar operational profiles in terms of patterns of analysis tool usage
- "Tool finder" suggests analysis tools based on other users' workflows including similar toolsets

Expert Finder

- Finds the user who has most often worked with the analysis tool in question
  - Can limit the search to one's social networks
  - Users can opt out of being considered experts
- Also can find geWorkbench "Power Users" who may be more familiar with the framework, rather than individual analysis tools

Software Engineering Considerations

- We benefited from geWorkbench's component-based architecture and its publish/subscribe model for transmitting events between components
- We were also able to easily integrate visualization features into the application because of geWorkbench's plugin architecture for user interface components

Current Status

- Logging of geWorkbench user activities will be included in next release
  - This will enable us to accumulate an initial knowledge store
- After any required clean-up, a small beta release of genSpace social networking features will be rolled out

Future Work

- Addressing “concept drift”
- Further investigation of impact on privacy, security, trust, etc. with respect to data sets and activity monitoring
- Tagging
- Just-in-time recommendations/suggestions

Conclusion

- We have presented an **approach** to knowledge sharing that is based on social networking metaphors
- We have also presented an **implementation** called genSpace, built on the geWorkbench platform for integrated genomics
- Potentially applicable to other kinds of scientists and engineers, including software engineers

genSpace: Exploring Social Networking Metaphors for Knowledge Sharing and Scientific Collaborative Work

Chris Murphy
cmurphy@cs.columbia.edu

enable (vt): *to make possible, practical, or easy*

PSL

PROGRAMMING SYSTEMS LAB
COLUMBIA UNIVERSITY

http://www.psl.cs.columbia.edu/