genSpace: Community Driven Knowledge Sharing in geWorkbench

Gail Kaiser’s Programming Systems Lab

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

Programming Systems Lab
Columbia University

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

Computer Science Department, Columbia University

June 1, 2012
Knowledge Sharing

- Scientists collaborating together in the same lab on the same project share:
  - Data: specimens, samples, materials, analyses
  - Tools: instruments, software, hardware
  - Knowledge: open discussion, whiteboard

- However, there are temporal (time) and physical (space) constraints

- This model does not scale to communities of scientists working on different projects but who could possibly learn from each other’s expertise, experience, etc.
CSCW Approaches

- Most current generation Computer-Supported Cooperative Work systems enable data sharing and/or tool sharing (e.g., EMSL Collaboratories, UIUC BioCoRE, Monod).
- However, these systems support relatively limited knowledge sharing - how/when/where/why to use tools and data.
- Knowledge sharing is partially enabled through labor intensive approaches: pubs, email lists, wikis, chat, shared display, etc. — may be outdated, requires active participation.
- We seek to enable automatic knowledge sharing — without requiring ”extra work” by scientists.
As geWorkbench continually expands with new analysis and visualization modules, choosing the right tools to use and learning how to chain these tools in sequence (workflows) can be very daunting, especially to novice users.

genSpace aims to alleviate this problem by providing recommendations about what to do first and what to do next.

Records, aggregates and data mines geWorkbench users’ activities with tools and tool sequences.

Users can ask questions such as what are the most frequently used tools and which workflows start with or contain a given tool, with the answers derived (and updated periodically) from the collective experience of geWorkbench users.
Our overall goal is to deliver community driven knowledge sharing, and to that end, we have implemented many features including:

- Tool and workflow recommendations
- Workflow sharing and visualization
- Community portal (website for tool and workflow comments)
- Social networking features (friends, networks, chat, profiles)
- Research notebook with automatic updates
Workflow Visualization

Viewing all workflows with a given tool
Workflow Visualization

Results
Recommendations

Workflow Visualization

Rating tools and identifying expert users
Workflow Suggestions

Default view

![Image of genSpace interface]

You haven’t used any tools! Next best rated tool to use: none.
Workflow Suggestions

Can be focused by tool
Workflow Suggestions

Updates as user performs analyses

Gail Kaiser's Programming Systems Lab

genSpace: Community Driven Knowledge Sharing in geWorkbench
Workflow Repository

Before accepting a shared workflow
Workflow Repository

After accepting a shared workflow
Workflow Repository

Viewing a workflow in the repository
Introduction to genSpace

genSpace Features

New this Year

Future Work

Community based knowledge

Workflow Statistics

Most Popular Tools

1. ARACNE Analysis
2. Hiearchical Clustering
3. T Test Analysis
4. Anova Analysis
5. MRA Analysis
6. HIEA Analysis
7. Gene Ontology Term Analysis (Ontologizer 2.0)
8. SOM Analysis
9. K-Means Clustering

Most Popular Tools at Start of Workflow

1. Hierarchical Clustering
2. T Test Analysis
3. ARACNE Analysis
4. Anova Analysis
5. Coefficient of Variation Filter
6. Fold Change Analysis
7. Entrez Gene ID Filter
8. SOM Analysis
9. PCA Analysis 3.0
10. MINDY Analysis

Most Popular Workflows

1. ARACNE Analysis, ARACNE Analysis
2. ARACNE Analysis
3. Hierarchical Clustering, Hierarchical Clustering
4. T Test Analysis, T Test Analysis
5. MRA Analysis
6. Hierarchical Clustering
7. Anova Analysis, Anova Analysis
8. ARACNE Analysis, ARACNE Analysis, ARACNE Analysis, ARACNE Analysis
9. Anova Analysis
10. SAM Analysis

Individual Tool Statistics

Anova Analysis

Total usage rate: 3735
Total usage rate at start of workflow: 743
The most popular tool used next to this tool: SkyBase Analysis
The most popular tool used before this tool: HouseKeeping Genes Normalizer
Social Networking

Viewing friends
Social Networking

Viewing profiles
Social Networking

Users by network and chat
Social Networking

Friend and network requests

![Friend and network requests screenshot]
Introduction to genSpace

community based knowledge

Social Networking

Privacy settings

![Image of genSpace interface showing privacy settings]

Gail Kaiser's Programming Systems Lab

genSpace: Community Driven Knowledge Sharing in geWorkbench
Introduction to genSpace

Community based knowledge

Community portal

Tool list

genSpace Features

New this Year

Future Work

Community portal

Tool list
Community based knowledge

Community portal

Tool detail
Community based knowledge

Community portal

Workflow detail

![Workflow Detail](image-url)
New this year

This year’s focus was on socially-oriented recommendations and collaboration tools.

- “People like you”
- Collaborative filtering and diversity in recommendations
- Research notebook

Additional notable tasks completed:

- Re-engineered genSpace recommendation platform to use Apache Mahout (scalable open source machine learning library)
- Developed prototype web-based genSpace user interface to integrate with the in-development web-based geWorkbench
“People like you”

Based on social connections and geWorkbench usage
Recommendations based on collaboration networks

- genSpace tool recommendations now take into account a user’s collaboration networks (including implicit “people like you” networks)
- This creates an issue of diversity:
  - Personalized recommendations can lead to a “bubble”
    - For example, if all members of a lab’s recommendations are based heavily on their lab mates, there will be no exposure to new ideas
  - genSpace solution: mix in results from other (visible) networks automatically, and allow users to specifically request more diverse results
Research Notebook

- Permanent archive of geWorkbench experiments
- Automatically logs every analysis, including dataset and parameters
- Annotations can optionally be added to any analysis
- History is fully searchable
Research Notebook

Viewing the notebook

![Image of genSpace interface]

Introduction to genSpace

- genSpace Features
  - New this Year
  - Future Work

Gail Kaiser’s Programming Systems Lab

Computer Science Department, Columbia University

genSpace: Community Driven Knowledge Sharing in geWorkbench
Research Notebook

Editing a notebook entry

![Image of genSpace interface]

- **Affy Detection Call Filter**
  - Date and time: 3/5/12 11:59 AM
  - Dataset: Bcell_100.exp
  - Comment: This didn't do what I thought it would. Need to try again.

- **Quantile Normalizer**
  - Date and time: 5/10/12 2:22 PM
  - Dataset: web100.exp
  - Comment: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec tincidunt gravida egestas.

- **Hierarchical Clustering**
  - Date and time: 5/10/12 2:19 PM
  - Dataset: brain_u.a5.296_filtered.exp
  - Comment: Duis convallis ligula at eros faucibus orare.

- **Standard Deviation Filter**
  - Date and time: 3/8/12 1:55 PM
  - Dataset: brain_u.a5.296_filtered.exp
  - Comment: This was my test.

- **Marker Centering Normalizer**
  - Date and time: 3/8/12 1:54 PM
  - Comment: Status: Ready
Research Notebook

Viewing parameters from an analysis
Research Notebook

Filtering view by tool

Gail Kaiser's Programming Systems Lab

Computer Science Department, Columbia University
Future Work

- Expand research notebook feature to allow groups of researchers to share notes
- Fully develop a web-based genSpace to complement the web-based geWorkbench
- Full support for workflows that include aborted and long-duration tools in geWorkbench
genSpace: Community Driven Knowledge Sharing in geWorkbench

Gail Kaiser’s Programming Systems Lab

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

Programming Systems Lab
Columbia University

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

Computer Science Department, Columbia University

June 1, 2012