Software User's Manual

Data Mining of Digital Library Usage Data

Team 07

Clients

Jewel Ward

Team Members

Project Manager: Maxim Krivokon
Developer: Bo Lee
Developer: Genesan Kim
Developer: Vu Nguyen
IV&Ver: Shing-Cheung Chan
IV&Ver: Marie Chi
IV&Ver: Kristine Guevara

Version History

Date	Author	Version	Changes made
4/5/2005	Genesan Kim	1.0	• Initial Version
4/10/2005	Genesan Kim	1.1	• Added section 2
4/26/2005	Vu Nguyen	1.5	• Updated Section 1.2 and 2.x
4/27/2005	Maks Krivokon	1.6	• Updated Sections 2 and 3

Table of Contents

Table of Tables

Table of Figures

1. Introduction

1.1 System Overview

The USC Digital Archive staff has requested an archive usage analysis system as a supplement to standard web metrics such as hit counts. Our implemented usage analysis system will take as input usage data from the digital library archive and will create relationship graphs for digital archive items. First a relationship report is generated thereafter fed into a graph clustering algorithm. This can then be opened with h3viewer where a 3D visualization is produced and can be analyzed properly. These graphs will give librarians a global structural view of the collections and will help them make decision on updating the collections. This system will be different from other systems because it automates creation of references/relations between digital library uses and documents, and can work with images and other items that do not have metadata or textual description.

1.2 System Requirements

1.2.1 Hardware

Computer:

- RAM: minimum 256 MB. Recommended: 512 MB
- HDD: minimum 100 MB disk space available. Recommended: 200MB.

1.2.2 Software

Operation Systems:

- Linux 2.6.9
- Mac OS X

2. Operational Procedures

Current system consists of two processing utilities: log2mcl, mcl2hv, and one visualization utility - h3viewer. Although all these parts have been packaged together, there are different commands to run each area. There is no initialization required for this software except that it is copied into the user's desired directory and on the computer of use.

2.1 Copy software to desired directory

All the system software components are contained in the folder 'datool'. In order to copy this into a user desired directory, run the following command in the OS shell:
cp –R datool ~
where ~ is your home directory, or can be replaced with any other desired directory.

* *From this point on, it will be assumed that the user runs all following commands in the software directory, specifically the 'datool' folder.*

2.2 Generate Object Relationships from Log File

Executable Name: log2mcl
User Function: Generate relationships
Input: Usage data log file. The log file is in text format generated by Digital Archive server.
Input Method: Pass in log filename along with command
Operational Result: .mci and .map files. .mci is the file representing object relationships. This file is used as an input for mcl2hv utility. *.map is the mapping to map object names with the numbers listed in .mci file.

Command:

./log2mcl/src/log2mcl <log_filename>

Where **<log_filename>** is the name of the log file that you wishes to analyze. This function terminates with the generation of the two files stated above.

2.3 Cluster the Relationships

Executable Name: mcl2hv
User Function: Cluster the relationship matrix
Input: Two files, .mci and a .map file that was generated by the previous function

Input Method: parameters passed to the utility using command line
Operational Result: log.h3v

Command:

./mcl2hv/src/mcl2hv <rep_filename> -I <g> -o <output_filename>

Where:

- <rep_filename>: is filename of relationship file generated by the log analysis utility (log2mcl).
- <g>: granularity level. Value is a decimal number from 1.2 to 5.0. Value 5.0 will tend to result in fine-grained clusters, and 1.2 will tend to result in very coarse grained clusters. The value recommended is 2.0
- <output_filename>: output file of the command. It is the plain text file following H3Viewer format.

2.4 View Clustered Tree

Executable Name: h3viewer
User Function: Generate 3D graph
Input: log.h3v, generated from the previous function
Input Method: Pass in filename on command line
Operational Result: New window with 3D graph

Step 1: Execute the command and load the GUI:

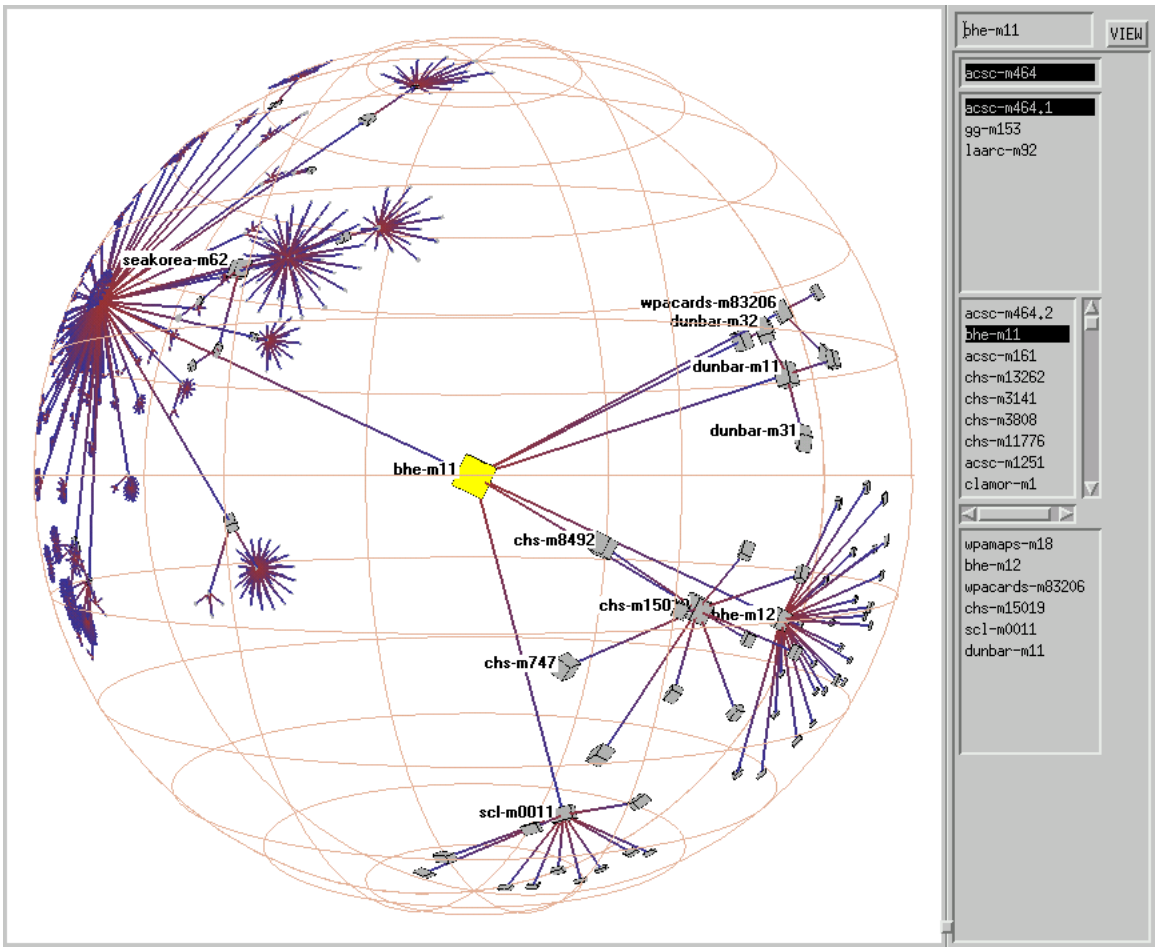
Command:

./h3viewer <h3viewer_filename>

Where **<h3viewer_filename>** specifies the file generated by mcl2hv utility above.

This function terminates by opening up a window that shows the 3D graph of nodes.

Step 2: View and Browse 3D Graph



The view is a 3d hyperbolic visualization of the object tree. Each node in the view represents an object from Digital Archive collection and edges represent hierarchical relationships parent-child relationships between objects.

The view of the tree is interactive and changes depending on user input:

Step 3a: Pan the view

Panning the view is done by pressing and holding left mouse button and dragging the mouse pointer in the 3d view area. Panning moves the focus in the view plane in the corresponding direction.

Step 3b: Rotate the view

Rotation of the view is done by pressing and holding right mouse button and dragging mouse pointer in the view area. Rotation is done around the focus point and allows bringing nodes and edges from periphery to the central part of the view, which allows examining them in detail.

Step 3c: Select object node

Selecting of the object node is done by clicking left mouse button in the view area on the corresponding box. Upon selection tree view is rotated and panned such that the selected node is brought into the view focus and tree is aligned such that all the ancestors of the selected node are on the left and all the descendants are on the right. The content of the right browsing panel is updated such that object lists reflect location in the tree hierarchy of the newly selected node.

The right pane is a file-system like browsing interface that allows user to see where the selected node belongs in the tree hierarchy. Each list contains all nodes on corresponding level of the tree. The list next to the last represents tree level of the currently selected node in the 3d view. The last level is the level of children for the selected node. All the lists above the next to the last represent levels of the tree where ancestors of the currently selected node reside. The following actions are available:

Step 3d: Select object from the list

User can select any item in the right pane browsing lists. The effects of such selection are identical to effects of action 3c.

Step 3: Search by object name

User can search for specific object by typing object name in the search text field and pressing enter. If the object with the specified name exists the system performs action 3c on the corresponding object. Otherwise nothing is done.

Step 4: View object thumbnail and metadata

User can view object's thumbnail and meta-data description in the separate web-browser windows by pressing button 'VIEW'. This will open a web-page of the Digital Archive collection with the selected object displayed.

3. Installation Procedures

Current system consists of two processing utilities: log2mcl, mcl2hv, and one visualization utility - h3viewer. The first two utilities are distributed in source code that should be compiled using GNU automake, autoconfig and make utilities. To compile a program, go into program directory (**datool/log2mcl, datool/mcl2hv or datool/h3viewer respectively**) and run the following command:

./autogen.sh && make

Compilation should produce an executable binary in./src subdirectory.

As an alternative to compiling the included visualization utility, a binary demo program with limited functionality can be obtained from http://graphics.stanford.edu/~munzner/h3/.

3.1 Initialization Procedures

None

3.2 Re-installation

Follow De-installation and Installation procedures.

3.3 De-installation

To de-install the system remove the directories where the system was extracted.

4. Troubleshooting

None

4.1 Frequently Asked Questions

None

4.2 Error Codes and Messages

No.	Message	Notes
1	Error: could not write to file: <filename>	This message occurs when the mcl2hv tool cannot write to the file <filename>.
2	Error: incorrect mapping. Total obj: <objectID>, cluster ID: <clustered>	The message is shown when the mapping at <objectID> and <clustered> is not correct
3	Not found object ID: <objectID>	The message is shown when the <objectID> is not found.
4	Cluster successfully! Output file: <ouput_file>	The message notifies users the mcl2hv tool completes clustering successfully.

5. Notes

None

6. Appendix