

Iteration Assessment Report

Data Mining of Digital Library Usage Data

Team 07

Maxim Krivokon – Project Manager

Bo Lee – Developer

Vu Nguyen – Developer

Genesan Kim – Developer

IV&Ver – Shing-Cheung Chan

IV&Ver – Marie Chi

IV&Ver – Kristine Guevara

April 24, 2005

Version History

Date	Author	Version	Changes made
4/24/2005	Vu Nguyen	1.0	• First Draft

Table of Contents

- VERSION HISTORY2**
- TABLE OF CONTENTS3**
- TABLE OF TABLES.....4**
 - 1. Overview 5
 - 1.1 References 5
 - 1.2 Capabilities Implemented 5
 - 1.3 Summary of Test Results 6
 - 1.4 Open Problems 7
 - 1.5 Objectives Reached 8
 - 2. Adherence to Plan 9
 - 3. Approach Used and Suggested Changes 10
 - 4. External Changes Occurred..... 11
 - 5. Suggested Actions 12

Table of Tables

Table 1 Capability implemented CAPI-1	5
Table 2 Capability implemented CAPI-2	6
Table 3 Capability implemented CAPI-3	6
Table 4 - Open Problems	7

1. Overview

This document details the assessment of the actual results of construction activity in the iteration #2 of Data Mining of Digital Library Usage Data project. The document will be presenting capabilities implemented and tested, evaluating deviations between planned and actual results of capabilities implementation and testing. This document also identifies outstanding issues, suggested changes and actions for improvement.

1.1 References

Spring 2005 CS577b Iteration #2 Plan

http://seacliff.usc.edu/~team7b/IOC/IP2_IOC_S05b_T07_V01.6.doc

Spring 2005 CS577b Life Cycle Plan

http://seacliff.usc.edu/~team7b/IOC/LCP_RLCA_S05b_T07_V06.0.doc

Spring 2005 CS577b Test Plan

http://seacliff.usc.edu/~team7b/IOC/TP_RLCA_S05b_T07.doc

Spring 2005 CS577b SSRD

http://seacliff.usc.edu/~team7b/IOC/SSRD_IOC_S05b_T07_V2.2.doc

Spring 2005 CS577b SSAD

http://seacliff.usc.edu/~team7b/IOC/SSAD_IOC_S05b_T07_V8.doc

1.2 Capabilities Implemented

The capabilities implemented during the iteration #2 are as follows:

Table 1 Capability implemented CAPI-6

CAPI-6	Browse analysis report
Relevant	C-03 Visualization log analysis results SR-6 Visualization SR-10 Do not visualize analysis report of corrupted format IR-5 Visualize analysis report LR-2 Usability – User-friendly interface for viewing item relationships and updating data LR-3 Usability – Maximizing the host resources LR-5 Performance – data of current scale
Deviation from	This capability is implemented in the iteration as a module in the

the plan	<p>system. The capability was fully implemented.</p> <p>In addition, the following three features that were added to the capability based on new requests from the client:</p> <ul style="list-style-type: none"> • Search: allows users to find an object in the H3Viewer • Browse selected object: users can browse the selected object by clicking on View button to open Browser window to view object online • The feature to show parents and siblings of the selected object/node
----------	---

The following capabilities were planned to be implemented in the iteration, but they were not implemented as the requirements have been changed.

Table 2 Capability not implemented CAPI-5

CAPI-5	Remove analysis report
Relevant	<p>C-02 Log Analysis</p> <p>IR-4 Manage analysis reports</p> <p>SR-11 Remove analysis report</p> <p>LR-1 System Dependability - Stable data import/export</p>
Deviation from the plan	The requirement has been changed and did not include this capability into the system. The analysis report is generated and stored as a file in file system. User can delete the report file using command line provided by the operating system.

Table 3 Capability not implemented CAPI-4

CAPI-4	Remove Usage Data
Relevant	<p>C-01 Manage Usage Log Data</p> <p>SR-2 Remove usage data imported from the specified log file</p> <p>IR-1 Manage usage data</p>
Deviation from the plan	The requirement has been changed and did not include this capability into the system. The usage data is generated and stored as a file in file system. User can delete the report file using command line provided by the operating system.

1.3 Summary of Test Results

The test items relevant to implemented capabilities were fully executed by developers. Based on the input data we have, all test items were passed. As required in the requirements, system must support up to 130,000 objects in the log data. The testing in

the iteration was focused on scalability and performance. The performance and scalability were fully tested using the data automatically generated by a program. We tested on the log data having 150,000 objects. The client was satisfied with the capabilities implemented.

Scalability Test Results:

Input size (entry)	0	1	100	500	1000	5000	10000	50000	100000	150000
Testing Code	S	S	S	S	S	S	S	S	S	S
Source Code	S	S	S	S	S	S	S	S	S	S
Result	P	P	P	P	P	P	P	P	P	P

Note: S: Successfully Run / N: Not successfully Run / P: Pass / F: Fail

Reliability Test Results:

Test cases	All same ids, in one big session	No relationship in 100* entry input	No relationship in 500* entry input
Testing Code	No relationship matrix generated	No relationship matrix generated	No relationship matrix generated
Source Code	No relationship matrix generated	No relationship matrix generated	No relationship matrix generated
Result	Pass	Pass	Pass

Note: (*) 100 and 500 are the numbers found on formal req.-based testing by IV&Vers

1.4 Open Problems

The following describes issues or limitations of the capabilities developed in the iteration:

Table 4 - Open Problems

Description	Design impact	Solution
To run the system, users have to type many	Medium	Consider combining command lines into shell

commands.		program. However, as commented by the client, she does not mind using command lines as soon as system provides stable capabilities.
The tree structure shown in the GUI is hard to follow and locate desired nodes	Medium. Only impact on the module to view the tree.	Consider using colors to denote the importance of nodes and relationships.

1.5 Objectives Reached

The goals stated in the iteration #2 plan were successfully reached. Besides required capabilities implemented, there are three new features added to the system. These features add significance to the success of the iteration. The client was very satisfied with these new features.

There are no major defects open for the next iteration. However, there are some limitations as mentioned in the Section 1.4. These limitations will be addressed in the next iteration of support stage.

Summary of objectives reached:

- The following use-cases implemented and their associated requirements (listed in Section 2.1) will be fully tested:
 - UC-06 Browse Analysis Report
- Scalability and performance:
 - The log analysis module supports up to 150,000 objects or nodes. Maximum time is less than 10 minutes (on recommended environment).
 - The clustering module supports up to 150,000 objects or nodes. Maximum time is less than 20 minutes (on recommended environment).
- Quality/Defect density
 - The system does not have any high and medium criticality defects
 - The system has fewer than 10 low criticality defects
- Other goals
 - All parts of the code follows the design

2. Adherence to Plan

The iteration #2 plan was executed successfully. As the project team has completed the capabilities earlier than planned, three additional features were developed in the iteration. The effective cooperation among team members is the main factor for achieving the plan.

3. Approach Used and Suggested Changes

The development team was divided into two groups, 2 developers each. Each group was responsible for implementing one module. Two developers of each group were working on the same set of functionality. The two developers of each group had intensive discussion to the same problems and came up with the best solutions. This practice helped to provide the good algorithms for each module as the algorithms significantly impact the performance and scalability of system. As two modules developed in the iteration are relatively independent, this approach did not result in any major breakage during integration. Overall, the concurrent development approach is suitable for the project, and it is recommended to keep this approach.

The team uses Subversion (a newer version of CVS) for controlling the project's artifacts including source code. The configuration process was implemented smoother than the previous iteration.

4. External Changes Occurred

The following are changes occurred in the iteration:

- Capabilities CAPI – 4 Remove Usage Data and CAPI – 5 Remove analysis report were not implemented in the iteration. To enable the flexibility for users, we proposed the use of files to store usage data/log data and analysis report. And the client accepted this change. Thus, the capabilities for removing these files are not necessary since users can delete files using operating system's commands.
- Three new features were added to the system:
 - Search: allows users to find an object in the H3Viewer
 - Browse selected object: users can browse the selected object by clicking on View button to open Browser window to view object online
 - The feature to show parents and siblings of the selected object/node

These features make the system easy to use and help users explore the tree more effectively. The change was agreed and made possible as the capabilities planned for the iteration were completed earlier.

5. Suggested Actions

This section discusses actions and recommended improvements for the project. Many of these actions or recommended improvements are from the client, users (ISD staff) and CSCI 577 teaching staff:

- Continue working on the system to improve the capabilities implemented
- Research and implement new ways for analyzing data based on the results generated by the system
- Cluster module needs updating to retrieve the most significant objects in the system
- Provide parameters for customizing the data analyzed. For example, the current data extracted from the log file is object ID and Session ID. It may be useful if the system allows users to select data they want to extract.