

ITERATION REPORT – 1

Iteration Number	Activities/Tasks/ Work Packages	Iteration Status	Done/Missing
1 st iteration	Machine Learning Algorithm Component	Passed: Addition of Weka jar files and use of Weka clustering and nearest neighbor libraries.	Nearest neighbor approach: Done Clustering approach: Partially Done
2 nd iteration	Evaluator Component	Pending	-
3 rd iteration	Distributed Systems Component	Pending	-
4 th iteration	Testing	Pending	-
5 th iteration	Web Service Package	Pending	-
5 th iteration	Bug Fixing	Pending	-
-	Final Product Package	Pending	-

The main task of the first iteration was the addition of Weka jar files to our Java project and the use of Weka clustering and nearest neighbor libraries. The aim was to improve our simple collaborative filtering algorithm. For this purpose, we added an attribute to our graph database; frequency. Frequency is the metric that shows how often a user listens to a specific song. In order to obtain this metric, we divided the number of listening action to the total time between the current time and the first time of action.

The integration of Weka API into the project was accomplished. We first used the

clustering library of Weka API and clustered the songs of similar users that are found to resemble the user of concern (the user that is to receive the recommendation). We chose the number of clusters according to the number of songs that should be clustered. However due to the problems of choosing the seed element for each cluster, we shifted to the nearest neighbor approach and compared the songs of the user of concern (only five songs with the highest frequencies) with the songs of similar users. We compared one song of the user with all songs of the similar users. Eventually, we selected the pair of songs with the least Euclidean distance. We achieved to get recommendations with the nearest neighbor approach. We are planning to get recommendations by using the clustering method as well. In this iteration clustering part remained unfinished. The related refactored codes were put into SVN repository.

The main task of the next iteration is the implementation of evaluation module. We will find out which of the two approaches (clustering and nearest neighbor) gives better results.