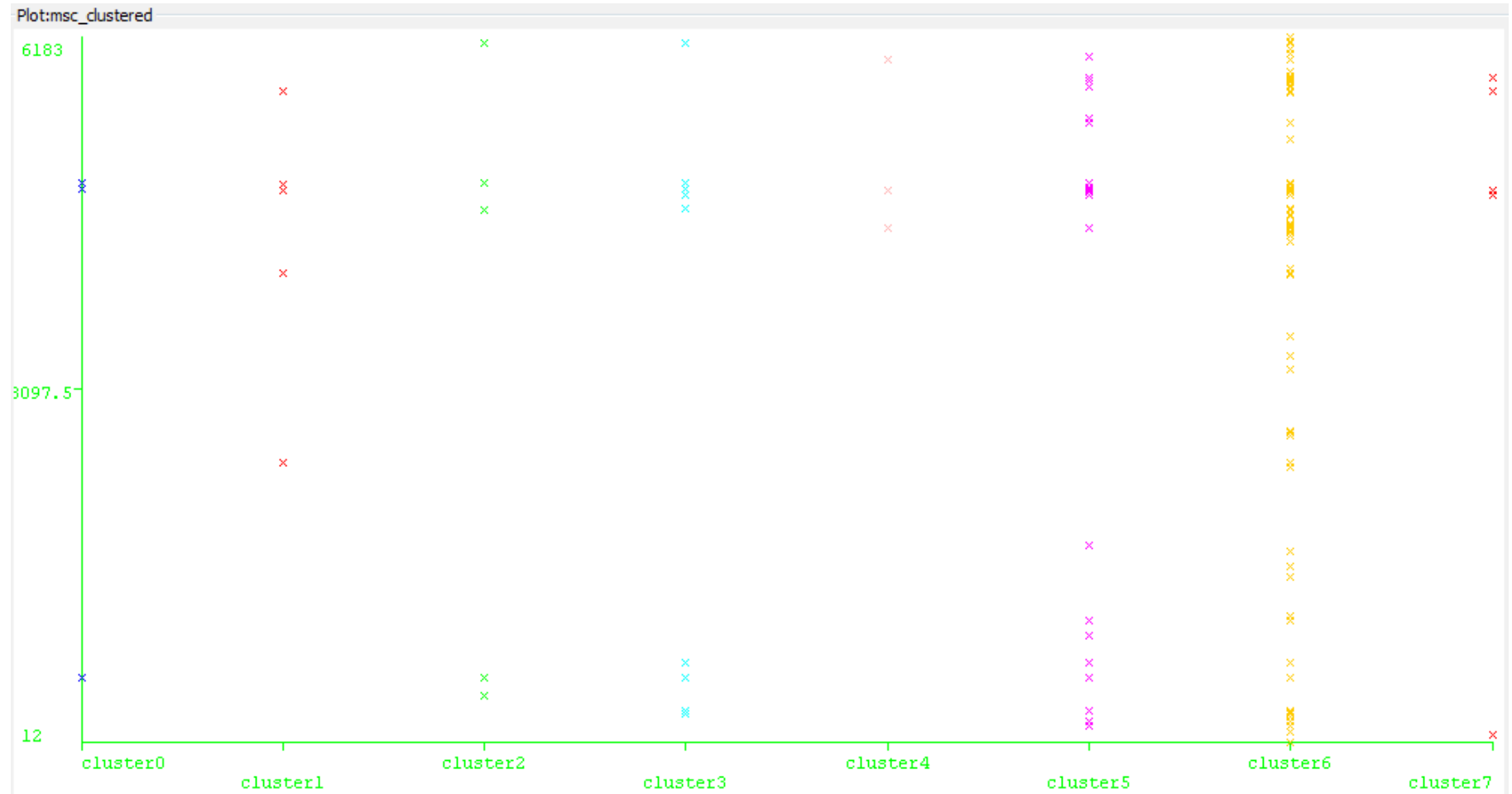


APPENDIX A – Expectation Maximisation Clustering of Training Set

The cluster plot shows training samples clustered into 7 clusters using default Weka EM parameters based on means and relative standard deviations of feature values. The y-axis indicates MSC class IDs to which each training instance is mapped. Additional information may be mined to form sub-clusters within these clusters and then learn the bag-of-words to MSC class mappings therein to improve accuracy.



APPENDIX B – System Architecture and Flow

