Processing of Raw Stock Data

I obtained the historical raw data for a year 2009/2010 from the site: <http://pages.swcp.com/stocks/>

I need to process this data before I can provide it as input to our clustering program. I filter closing prices of last day of every month from one year historical data. I executed the following steps to come up with an 11-point representation of a stock.

* + I generated returns for each stock using r(t+1) =  (s(t+1) - s(t))/s(t) generating 11 data points for each stock
  + Returns Ire normalized using (r(t) - mu)/std where “mu” is mean and std is standard deviation

Please note that the file stocks.txt does not have all the stocks in S&P500. I had to filter out certain stocks as the data was not complete for those stocks.