ECE568

WEB Application

Design report of collecting data module

Chaoji Zuo(cz296)

# Overall design

In our project, we choose python as our major programming language because it can easily produce data and do the machine learning.

We choose to use MongoDB as our database because this database is friendly to stock the array-like data like stock data. Besides, we use pymongo to connect MongoDB in python.

We also use the famous python data analysis library NumPy and Pandas to produces data.

# Data collecting part

We have a class names get\_stock to achieve data collecting function and data store function.

* External APIs

To acquire the stock data, we used Alpha Vantage API and a python package names yahoofinancials2.

We used Alpha Vantage to gain the real time data and yahoofinancials2 for the historical data.

* Interior interfaces

Inside the class get\_stock\_data, the data collecting part would provide two interface: get\_historical\_data(), get\_realtime\_data(). Both of the interface would return two list, one for build the csv, one for store in the database.

get\_historical\_data would return the historical data of one year with the interval is weekly, contains time, open, close, high, low and volume.

get\_realtime\_data would return the intraday data with the interval is a minute, contains price, time and volume.

# Data storing part

After collecting data from the internet, we have to store our data to our local MongoDB database and save as a csv file.

* MongoDB

We use pymongo package in python to connect our database. We use pymongo. MongoClient to connect our client, and use myclient['stockdb'] to get into our database, for each ticker, we would create two collection by mydb[ticker] and mydb[ticker + '\_realtime'], which store historical data and realtime data.

* CSV

When we get the data in previous part, we would take the type of data into list and save the data into csv file. We would create (ticker+’.csv’) in our get\_historical\_data() to store historical data and create (ticker+’\_realtime.csv’) in our get\_realtime\_data() to store realtime data.