

# Independent Project Proposal

## Summer 2009

Michael Rechenthin  
Loyola University Chicago, Chicago, IL 60640  
mrechenthin@gmail.com

**Abstract**—For my summer independent project, I propose writing a program that retrieves historical quotes for all 500 stocks in the S&P 500 index and store the data in a database. Furthermore, the program will allow user interaction. When a stock is entered by the user, the database is queried which results in computation of the top five most highly correlated stocks over specific time frames.

### I. INTRODUCTION

As a trader on the floor of the Chicago Stock Exchange for eight years, my interest lie in equity markets and the development of trading systems. To analyze and mine the market data first requires a method to efficiently retrieve and store the data. This paper will propose how I intend to develop this method and will also demonstrate why finding strongly correlated equities is a useful tool for traders.

More specifically, my three main objectives for this summer project include:

- Create a program that retrives historical market data for all 500 stocks in the S&P 500<sup>1</sup>
- Store the data in a database
- Data mine for stocks that best correlate with a user inputed stock symbol

Futuremore, I will be starting a doctoral program in Management Science at the University of Iowa in the Fall, and I plan to use the program I develop to assist me in my future research by giving me easier access to market data.

### II. HOW CORRELATIONS ARE USEFUL TO TRADERS

There is an aphorism in trading, “*All boats float or sink with the tide.*” This describes how individual stocks rise and fall depending on how the overall market is doing. Individual stocks are often closely correlated to other individual stocks; therefore, short-term traders find it useful to look at these correlations.

For example, IBM (International Business Machines Corp), by the nature of its business, is closely related to ACN (Accenture Ltd.), MSFT (Microsoft Corp.), and HPQ (Hewlett-Packard Company). Using Pearsons correlation coefficient<sup>2</sup> on

<sup>1</sup>Standard & Poor’s 500 is an index of large-cap American stocks that are traded on either the New York Stock Exchange or NASDAQ. The index is maintained by Standard & Poor’s, which is a division of McGraw-Hill.

<sup>2</sup>An excellent site for a futher understanding of correlation coefficient is: <http://mathworld.wolfram.com/CorrelationCoefficient.html>

the closing prices for the last 500 trading days of data<sup>3</sup> using SAS[3], we can see the following:

|      |      |
|------|------|
|      | IBM  |
| IBM  | 1.00 |
| ACN  | .739 |
| MSFT | .565 |
| HPQ  | .716 |

In the above example, although the correlation between IBM and ACN is high, it does not imply causality[2]. Even though this measure is not a forward-looking measurement, and we cannot say that the movement in stock ACN impacts the movement in stock IBM, there are useful applications with this data for an experienced trader. One such use is for arbitrage. The Efficient Market Theory claims that prices contain all available information at a given instant of time. If the arbitrageur pocesses some information not reflected in the price, then a profit could be made, and doing so would bring the price closer to its true value[6].

### III. WHAT PROGRAMS OR SITES CURRENTLY PROVIDE THIS DATA?

I have not found any programs or sites that currently provide the type of data I am seeking. There are several sites that provide the correlation between two stock symbols; however, my project would greatly expand upon this. Specifically, I want to find the top five correlations with the user inputed stock. These are correlations that may not be completely obvious, but would be very useful in trading.

### IV. MARKET DATA

#### A. Requirements

For this project, I will need a provider that:

- provides historical daily data over the past 100 days
- provides real-time intraday data on a minute-by-minute time period
- allow for the retrieval of all 500 stocks in the S&P 500 index
- has application programming interface (API) availability
- is inexpensive

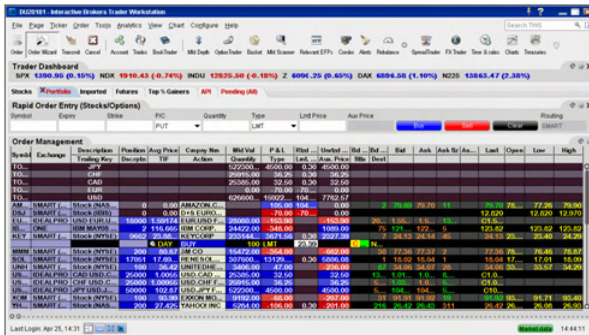
<sup>3</sup>May 12, 2009 to May 18, 2007

## B. Data Provider - Interactive Brokers

The data provider that I selected for this project is Interactive Broker's Trader Workstation. It fulfills all of my requirements. The cost is low, \$10 a month, and Interactive Brokers supplies an API. Furthermore, I have an account already setup, which is convenient.

Interactive Brokers has an online trading platform named Trader Workstation, or TWS, that allows trading and managing orders for all types of financial products[4]. To be able to retrieve historical information via the API, the TWS must be logged in and running in the background. The TWS can be seen in Figure 1.

Figure 1. Interactive Broker's TWS



A particular interesting method within the API is `reqHistoricalData`, which retrieves historical market data[5]:

```
void reqHistoricalData (
    int id,
    Contract contract,
    String endTime,
    String durationStr,
    String barSizeSetting,
    String whatToShow,
    int useRTH,
    int formatDate
)
```

## V. DATA STORAGE

For this project, I will use a relational database, specifically MySQL[7]. An example of the tables are in Figure 2 and Figure 3.

Figure 2. Proposed daily data table

| Key  | Symbol | Date       | Num | Last Sale |
|------|--------|------------|-----|-----------|
| IBM1 | IBM    | 2009-05-12 | 1   | 103.94    |
| IBM2 | IBM    | 2009-05-11 | 2   | 102.90    |
| IBM3 | IBM    | 2009-05-10 | 3   | 101.49    |

Figure 3. Proposed intraday data table

| Key  | Symbol | Date       | Time  | Num | Last Sale |
|------|--------|------------|-------|-----|-----------|
| IBM1 | IBM    | 2009-05-12 | 09:30 | 1   | 103.94    |
| IBM2 | IBM    | 2009-05-12 | 09:29 | 2   | 103.88    |
| IBM3 | IBM    | 2009-05-12 | 09:28 | 3   | 103.72    |

## VI. STATISTICAL PACKAGE

For the data analysis, I will use the Apache Common's Mathematical Library[8]. The class location is shown below:

```
org.apache.commons.math.stat.-
correlation.PearsonsCorrelation
```

This class has full *javadoc* documentation and it contains multiple constructor support.

Pearsons correlation coefficient is shown below[1]:

$$r_{xy} = \frac{\sum x_i y_i - (\sum x_i \sum y_i) / n}{\sqrt{\sum x_i^2 - (\sum x_i)^2 / n} \sqrt{\sum y_i^2 - (\sum y_i)^2 / n}}$$

## VII. PROJECT MANAGEMENT

A proper breakdown of the work is required to estimate time more efficiently and ensure that the project is completed by the deadline. I broke the work into work packages, activities, and tasks. My Work Breakdown Structure and Gantt chart can be found in the Appendix.

## VIII. CONCLUSION

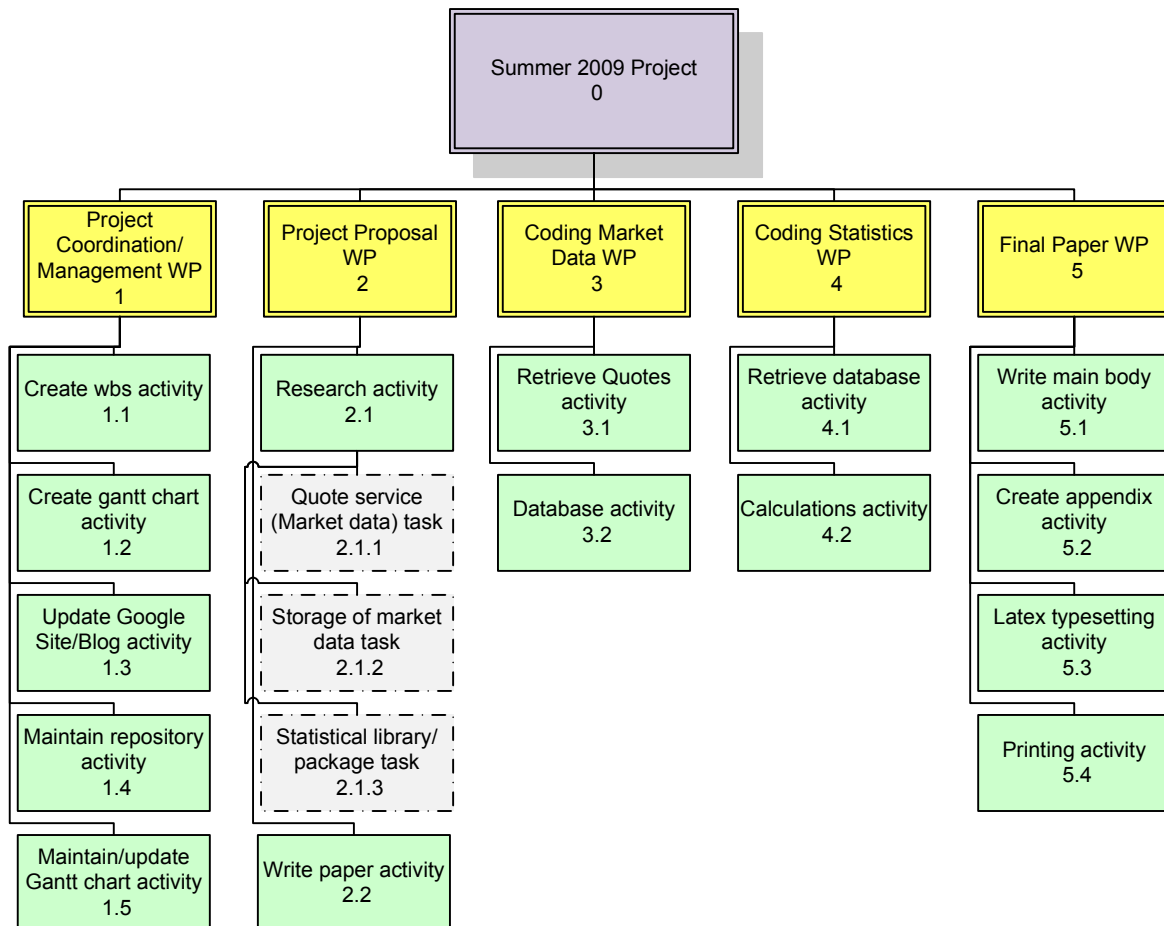
It is exciting to know that the idea of finding correlation could expand beyond equities. For example, one could use correlations to investigate the relationship of oil with particular stocks within the oil sector. Or, one could investigate the correlations between currencies and American Depositary Receipts (ADRs)<sup>4</sup> of a particular country. I believe this project is not only very interesting, but also very useful for my future research needs and beyond. I am also confident that I can complete it in the time allowed.

## REFERENCES

- [1] David R. Anderson, Dennis J. Sweeney, and Thomas A. Williams. *Statistics For Business and Economics*. West Publishing Company, 6th, edition, 1996.
- [2] Ronald P. Cody and Jeffrey K. Smith. *Applied Statistics and the SAS Programming Language*. Pearson Prentice Hall, 5th, edition, 2006.
- [3] SAS Institute. SAS. <http://www.sas.com/>.
- [4] Interactive Brokers LLC. Getting Started with the TWS Java API, August 2008.
- [5] Interactive Brokers LLC. API Reference Guide, February 2009.
- [6] Matthieu Wyart and Jean-Philippe Bouchaud. Self-referential behaviour, overreaction and conventions in financial markets. <http://ssrn.com/abstract=391861>.
- [7] MySQL AB/Sun Microsystems. MySQL. <http://mysql.com>.
- [8] The Apache Software Foundation. The Apache Commons Mathematics Library. <http://commons.apache.org/math/>.

<sup>4</sup>ADRs enable US investors to buy shares in foreign companies. ADRs trade on US exchanges similar to regular stocks.

## Work Breakdown Schedule



### WBS Dictionary

|          |   |   |
|----------|---|---|
| <b>0</b> | <b>Summer 2009 Project</b>                |   |
| <b>1</b> | <b>Project Coordination/Management WP</b> |   |
| 1.1      | Create wbs activity                       | create the WBS via Microsoft Visio                  |
| 1.2      | Create gantt chart activity               | create the Gantt chart via Microsoft Project        |
| 1.3      | Update Google Site/Blog activity          | keep the Google Site updated                        |
| 1.4      | Maintain repository activity              | subversion via Netbeans                             |
| 1.5      | Maintain/update Gantt chart activity      | maintenance of Gantt chart/PM via Microsoft Project |
| <b>2</b> | <b>Project Proposal WP</b>                |   |
| 2.1      | Research activity                         | research via web/books/library                      |
| 2.1.1    | Quote service (Market data) task          | determine most suitable quote service               |
| 2.1.2    | Storage of market data task               | determine how to store the quote/market data        |
| 2.1.3    | Statistical library/package task          | determine which statistical package to use...if any |
| 2.2      | Write paper activity                      | write the proposal of said research                 |
| <b>3</b> | <b>Coding Market Data WP</b>              |   |
| 3.1      | Retrieve Quotes activity                  | retrieve quotes from market service                 |
| 3.2      | Database activity                         | store quotes in suitable database                   |
| <b>4</b> | <b>Coding Statistics WP</b>               |   |
| 4.1      | Retrieve database activity                | retrieve quotes from database                       |
| 4.2      | Calculations activity                     | statistical calculations                            |
| <b>5</b> | <b>Final Paper WP</b>                     |   |
| 5.1      | Write main body activity                  | write the main body of the paper                    |
| 5.2      | Create appendix activity                  | create the appendix                                 |
| 5.3      | Latex typesetting activity                | tie up loose ends with latex                        |
| 5.4      | Printing activity                         | print and turn in                                   |

