

**Diploma in Financial Business Informatics**

Major Project (CMP3801)

**AY2012/2013 Apr Semester Term B**



***Submission of Major Project Report***

Project Title: nTrader

Supervisor: Esther Chia

Evaluator: Wong Kok Keong

|  |  |
| --- | --- |
| **Student name** | **Admission number** |
| Soh Jun Jie | 1003346B |
| Ong Qi Yong | 1004142D |
| Leung Kai Yiu | 1001423A |
| Yan Dawei | 1001433J |

**ACKNOWLEDGEMENTS & DISCLAIMER**

The data and Application Programming Interfaces (API) used as part of this project are sourced from the following company websites:







Specifically, the group used the financial news data, and the market price data API from Bloomberg. Financial terms/jargon data are sourced from Financial Times Lexicon. The Geomap charting tools used by the group in creating the map chart is based off from the Google Geomap API, and the stock charting tool is based off from HighCharts JS stock charting API. Economic indicators data are sourced directly from The World Bank.

The used of the data /API from these extracted company sources are used only for the purpose of MP project and academic purposes i.e. personal use, therefore the group shall not be held liable for the developed website. Links represented in the website also does not constitute any form of endorsement to any information of those linked sites. Out of courtesy and respect, the group had also contacted the relevant parties in the use of their data or API. However there were also instances where the group fails to contact the relevant parties. For that, the group wished to express their deepest regrets and continue to find means to contact these parties.

The group also wished to acknowledge the project supervisor: Ms Esther Chia and project evaluator: Mr Wong Kok Keong, for their valuable assistance and feedback throughout the course of the project.

Table of Contents

[1. INTRODUCTION 1](#_Toc346133466)

[1.1. Background 1](#_Toc346133467)

[1.2. Project goals 1](#_Toc346133468)

[1.3. nTrader functionalities & products’ purposes 2](#_Toc346133469)

[1.3.1. nConsole 2](#_Toc346133470)

[1.3.2. nIntel 3](#_Toc346133471)

[1.3.3. WordLearn HQ 3](#_Toc346133472)

[1.3.4. Amazon EC2 server’s role in supporting nTrader 4](#_Toc346133473)

[1.3.5. MYSQL Database 4](#_Toc346133474)

[1.3.6. Alternative PHP Caching (APC) 4](#_Toc346133475)

[1.4. Scope 5](#_Toc346133476)

[1.4.1. Project scope 5](#_Toc346133477)

[1.4.2. Project constraints & limitations 5](#_Toc346133478)

[1.5. Methodology 6](#_Toc346133479)

[2. FINDINGS OF LITERATURE REVIEW 6](#_Toc346133480)

[2.1. Natural Language Processing (NLP) in nTrader 6](#_Toc346133481)

[2.2. Database and Web structure 7](#_Toc346133482)

[2.3. Analysis on choice of News Provider 8](#_Toc346133483)

[2.4. Algorithm to Derive news sentiment 9](#_Toc346133484)

[3. PROBLEM RECOGNITION 9](#_Toc346133485)

[4. REQUIREMENTS AND PROPOSED SOLUTION PACKAGING 10](#_Toc346133486)

[4.1. Analysis of user requirements 10](#_Toc346133487)

[4.2. Project solution and feasibility 11](#_Toc346133488)

[4.2.1. Solution design 11](#_Toc346133489)

[4.2.2. Database security concerns and design 18](#_Toc346133490)

[5. FEASIBILITY STUDY 20](#_Toc346133491)

[5.1. Technical feasibility 20](#_Toc346133492)

[5.2. Economic feasibility 22](#_Toc346133493)

[5.3. Legal feasibility 23](#_Toc346133494)

[5.4. Operational feasibility 24](#_Toc346133495)

[6. PROJECT TRIALS & TESTING 24](#_Toc346133496)

[6.1. Difficulties faced 24](#_Toc346133497)

[6.1.1. Graphical User Interface 24](#_Toc346133498)

[6.1.2. Data gathering 26](#_Toc346133499)

[6.1.3. Change Control Management 28](#_Toc346133500)

[7. EVALUATION & USER FEEDBACK 29](#_Toc346133501)

[7.1. Evaluation of project processes 29](#_Toc346133502)

[7.2. User evaluation 31](#_Toc346133503)

[7.3. Self-evaluation of Project nTrader 33](#_Toc346133504)

[7.3.1. Strengths 33](#_Toc346133505)

[7.3.2. Weaknesses 33](#_Toc346133506)

[8. RECOMMENDATIONS 34](#_Toc346133507)

[8.1. Technical improvements 34](#_Toc346133508)

[8.1.1. Security 34](#_Toc346133509)

[8.1.2. Cloud server management 35](#_Toc346133510)

[8.1.3. Natural Language Processing (NLP) 35](#_Toc346133511)

[8.1.4. Browser compatibility 36](#_Toc346133512)

[8.1.5. Screen resolution 36](#_Toc346133513)

[8.2. Non-Technical Improvements 37](#_Toc346133514)

[8.2.1. Gaming 37](#_Toc346133515)

[8.2.2. Personalization 37](#_Toc346133516)

[8.2.3. Text-based support 37](#_Toc346133517)

[8.2.4. Multi-Language support 38](#_Toc346133518)

[8.2.5. SMS support 38](#_Toc346133519)

[8.2.6. WordLearn HQ Improvement 39](#_Toc346133520)

[8.2.7. Trading 39](#_Toc346133521)

[8.2.8. Forums 39](#_Toc346133522)

[9. CONCLUSIONS 40](#_Toc346133523)

[ILLUSTRATIONS 41](#_Toc346133524)

[BIBLIOGRAPHY 55](#_Toc346133525)

**LIST OF ILLUSTRATIONS**

|  |  |  |
| --- | --- | --- |
| No. | Description | Page |
| Figure 1 | Simple random sampling of 30 news headlines with the word ‘stocks’ in it | 41 |
| Figure 2 | Accuracy statistic of bull-bear sentiment generator algorithm | 42 |
| Figure 3 | Literacy scores for tier segments | 42 |
| Figure 4 | Group’s database logical data model | 43 |
| Figure 5 | Extracted data using Excel VBA data extraction algorithm | 43 |
| Figure 6 | Cost of Amazon EC2 Cloud Server | 44 |
| Figure 7 | Bloomberg’s robots.txt | 45 |
| Figure 8 | Financial Times Lexicon’s robots.txt | 45 |
| Figure 9 | Bloomberg’s permission request form for use of site materials | 46 |
| Figure 10 | Secured Shell console image | 47 |
| Figure 11 | Secured Shell File Transfer Console | 48 |
| Figure 12 | Excel side application web crawler | 49 |
| Figure 13 | SQL insert statement batch generator | 49 |
| Figure 14 | Initial Change Control framework | 50 |
| Figure 15 | Updated Change Control framework | 50 |
| Figure 16 | nTrader products & services survey | 51 |
| Figure 17 | Usefulness of nIndicate feature | 52 |
| Figure 18 | Usefulness of nIntel charting feature | 52 |
| Figure 19 | Usefulness of WordLearn HQ educational portal | 53 |
| Figure 20 | Usefulness of nTranslate feature | 53 |
| Figure 21 | Most useful feature(s) of nTrade | 54 |

**APPENDICES LIST**

A1 – Terms of Reference

A2 – Project schedule

* Planned schedule
* Actual schedule

A3 – Workload distribution

A4 – System design

* System architecture
* Database structure
* Logical database model

A5 – Fact Sheet

* Product Survey Form
* Product Survey Results

A6 – FILA Chart

A7 – Minutes of Meeting

A8 – Weekly Reflection

* Soh Jun Jie – 1003346B
* Ong Qi Yong – 1004142D
* Leung Kai Yiu – 1001423A
* Yan Dawei – 1001433J

A9 – Literature Review

* Natural Language Processing Methodology
* Technology
* News Provider Comparison
* News Analytics Applications in the Financial Industry

A10 – Post-Implementation Reflection

**ABSTRACT**

Financial news is an important medium for investors to make decisions for their investments. For beginner investors, many of them do not have the luxury of time to read financial news or do not have the financial literacy to understand news which affects their decision making. The purpose of this report is to present a solution known as “nTrader” which will help facilitate news understanding and help investors make wiser investment decision. nTrader is a web based project that consist of nConsole which is an interactive platform to aid investors understanding of financial news. It comprises of four main components which help by translating news, analyzing news, perform business analytics function and educating its users. nTrader is a ten weeks project where the project team aim to use various technologies both externally available and internally developed, to create a powerful engine that is capable of executing natural language processing algorithms that can provide financial signals and insights to its users to be used for trading. However due to the various difficulties and limitations the project team face, nTrader was not able to cover the trading aspects of the project and instead cover a greater amount of learning and analysis content. Although nTrader has served its intended purpose well, user feedback has displayed the difference in mindset between financial literate users and non-financial literate users when using the nConsole. Even though the nTrader project is completed, there are still many room of improvement for the project such as security, natural language processing engine and education features.

# INTRODUCTION

## Background

The objective of this project report is to document work done related to the “nTrader” project. The initiation of the project is triggered by a key problem in the context of investment decision-making for retail level investors whom are just beginning to start investing. The problem is the unbridgeable technology gap between institutional investors and beginner investors which often results in these investors not having access to tools that helps derive effective investment decision. In addition, these investors do not have the luxury of time to comprehend financial news much less being financial literate to make effective investment decisions.

In summary, beginner investors have the following characteristics that warrant our attention.

* Do not have the luxury of time to read news
* Do not understand financial terms and jargons
* Do not have the capacity to digest huge amount of information
* Do not know how to derive investment decisions through text and qualitative data

In order to provide these investors with access to adequate investment analysis technology that will help them get started with investment decisions, the group developed the nTrader web application project. The main aim of the web application is to resolve the technical aspects of financial jargons within financial news, and give users a machine-generated sentiment signal base on the nature of the financial news for their considerations.

## Project goals

The goal of the nTrader web application project is to enable beginner investors to effectively appreciate the impact of financial news to financial markets. In attempting this, the project will interpret financial jargons in financial news for users while providing analytical charting tools used in conjunction with the group’s machine-generated sentiment indicators for news. Beginner investors will be able to appreciate financial news better with financial jargons interpreted as well as derive investment decision-making base on said charting of news sentiment indicators.

## nTrader functionalities & products’ purposes

The following describes the products covered by nTrader, their functionalities and description, and the background systems integral to nTrader.

### nConsole

The purpose of nConsole is to allow users the ability to view machine-generated sentiment signals for the news displayed by the nTrader web application. nConsole is able to stream to users the financial news supported by nTrader, at an on-demand basis. The following are the other purposes of the nConsole product.

* Extraction of financial news from relevant news providers along with the display of machine-generated sentiment signal which rate the nature of the extracted financial news. The group used their own text parsing algorithm to derive sentiment signal.
* Simple search and advance search functions in nConsole allow users to find or filter financial news base on keywords that can search within news content, type of news (e.g. stocks, commodities, bonds), by country which the news is related to, date range, or sentiment indicator.
* nConsole is able to perform on-demand streaming of financial news content. Within the extracted news content, nConsole is able to identify key events, giving users an overall summary of the news and its highlights.
* nConsole Jargon Translation feature is able to display the definition of financial jargons found in the extracted financial news. On click the financial jargon, users can view its definition. User can also “favourite” financial jargons in their account settings and review the financial jargons at a later time.

### nIntel

The purpose of nIntel is to translate nConsole’s machine-generated news sentiment signal into quantitative data suitable for technical analysis and investment decision-making. nIntel has 2 components: Graph, and Map. The following are the functionalities of nIntel components.

* nIntel Graph feature allows users to perform basic technical analysis on supported financial instruments. Supported technical indicators are SMA (Simple Moving Average), RSI (Relative Strength Index), and Stochastic. The Graph is also able to display news event flags which are bound with sentiment rating so that news can be taken as part of users’ trading strategy.
* nIntel Map feature allows users to visualize news in the form of a map chart. The Map plots sentiment rating of countries which are machine-generated from financial news the nConsole function extracted. The Map can also be manipulated by a date range input. Users of the Map feature can perform basic E-I-C analysis for stock picking.
* Within nIntel Map feature, sentiment rating charting and display of economic indicator are also supported. On click of a country in the nIntel Map feature, nIntel will be able to display analytical chart for country sentiment ratings and the country’s economic indicators.

### WordLearn HQ

The purpose of WordLearn HQ is to aggregate the learning the process of financial terms in a single area. The aim of this function is to improve financial literacy of users through the introduction of word games, allowing searching of financial terms to find its definition, term of the day feature, and management of “favourited” financial jargons through the nConsole function. The group has developed 2 word games namely “Find The Jargon” and “Guess The Jargon”. The following are the functionalities of the components of WordLearn HQ.

* The “Find The Jargon” word game requires users to identify financial jargons in a randomly generated financial news passage. The aim of the game is to encourage users to read financial news while playing the game.
* The “Guess The Jargon word game simulate the traditional hangman word game where users are required to guess the financial jargon word when given its definition.
* The “Term of The Day” is an event scheduled daily to display a random financial jargon.
* The Jargon Management function in WordLearn HQ allows user to manage jargons they have favourite e.g. Remove favourite jargons.

### Amazon EC2 server’s role in supporting nTrader

In order to reduce manual data maintenance, the group employs the use of EC2 server’s crontab feature to schedule archiving of Bloomberg financial news on a daily basis. In addition, EC2 server also acts as a backup for the restoration of project files with its AMI (Amazon Machine Images) imaging service in the event source codes deleted by accident. The group also used the “mail” feature that is internal to the server to generate password recovery email to users.

### MYSQL Database

The group deploys the EC2 MYSQL database to support operation of nTrader website. Extending the use of database more than simply storage of analytical and persistent data, the group used the database to schedule execution of SQL event. This is used for example in our “Term of The Day” function to insert a random financial jargon into a table daily. The group also used the “routine” feature of MYSQL to protect against SQL injection by having SQL statement to only accept only a given list of parameters.

### Alternative PHP Caching (APC)

The main purpose of APC is to reduce the number of connections with the database in retrieving possibly static variables or variables that are infrequently changed. The group deploys APC caching solution to compensate for PHP limited processing speed and allow webpages to load faster, hence improving user experience.

## Scope

### Project scope

Due to limited manpower resources and time constraint of 10 weeks, the group does not guarantees a fully functional nTrader web application. In order to demonstrate a working prototype web application system and due to other considerations, nTrader adopted the following project scope.

* Financial news extraction and archive methodology for the web application will be Bloomberg-based. The rationale for this is due to Bloomberg being a reliable financial news provider and its news being properly formatted and standardized for machine readability.
* Only news article about countries and having the word ‘stocks’ as part of its headline will be extracted from Bloomberg for analysis by the web application. Refer to Illustration: Figure 1, the approach taken is due to news having ‘stocks’ as part of its headline conforms to a specific format for machine readable.
* The main source for financial jargons and definition data will come from Financial Times Lexicon website. The main rationale is due to the site keeping a record of recent changes to their lexicon database. For future scalability, the group web application can perform changes to its jargon database base on changes on the source website.
* The group will be limiting the full extent of nTrader web application functionality to only Google Chrome and Fire Fox browser as both have rich support for JQuery programming framework to deliver rich animation content. Internet explorer browser will not be supported at all as these animation content are not supported. The nTrader web application is best viewed in Google Chrome as majority of the design, testing and development are through Google Chrome browser.

### Project constraints & limitations

1. The schedule for the Major Project is 10 weeks from 5-Nov-2012 to 11-Nov-2012.
2. The group is expected to complete the project under a zero-cost budget constraint.
3. In terms of manpower, the group is constraint to 4 members. For the purpose of news analytics algorithm to derive sentiment rating for financial news, the lack of qualified linguists and trained programmers knowledgeable in the implementation of Natural Language Processing (NLP) algorithm are constraints.
4. The members of the group are knowledgeable in programming skills limited to only PHP, Javascript, CSS, SQL, and Excel VBA.
5. For technological constraints, the usage of Temasek Polytechnic WIFI service causes blockage of certain server ports. Standard lab computers also do not have professional web designing programs (e.g. Adobe Dreamweaver) and debugging tools preinstalled.

## Methodology

Information in this report was obtained from nTrader product/service feedback, literature review, websites legal terms of services, and statistical report by Monetary Authority of Singapore.

# FINDINGS OF LITERATURE REVIEW

## Natural Language Processing (NLP) in nTrader

nTrader web application system relies on the group’s traditional text parsing algorithm to identify the nature of financial news and assign news sentiment rating. While the error rate of our algorithm is less than 8.5% (refer to Illustration: figure 2), it does not follow according to linguistics rules of the English language. This is possible as news headlines contain little linguistics complexity versus complex language structure within news content. In order to provide in-depth analysis of news content instead, the group researched on NLP algorithms, which strictly follows the linguistics rules by identifying the function of words and phrases in a given sentence (part-of-speech), thereby removing lexical ambiguity. The group had implemented a miniature element of NLP algorithm into nTrader. This NLP algorithm created by the group is able to search for relevant financial entities within a given financial news and capture related events and the entity’s status if available within the news.

The group had also investigated briefly in machine-learning algorithms which can potentially remove the need for continuous maintenance of programming codes. This will be possible by creating a system that improves its knowledge-based base when exposed to new inputs. Refer to *literature review on Natural Language Processing page 9*, the group intent to create machine-learning algorithm that will learn new financial jargons as they are created.

## Database and Web structure

nTrader project deals with a high amount of text, news and financial information which has to be stored in a proper lexicon database. However our group has found many types of lexicon databases through research and each of these lexicon databases has their own strengths and weaknesses. Through intensive research on the various types of database, the group has identified a database structure known as “MorDebe database” to be highly suitable for the nTrader project. The “MorDebe database” uses a structure where full-form words are stored and used to reference its other inflected paradigms. Through this design, the group could easily tag new information to full-form words and create various relationships between words without having to change the database design greatly.

However a “MorDebe database” is difficult to build since the number of full-form words far exceeds the number of lemma in a language and thus it would take an extremely long period of time to create a functional “MorDebe database” that is able to not only store full-form words but also able to allow the group to use it effectively for word relationship tagging. In order to reduce the time taken to set up a functional database, the group has opted to use an existing lexicon database known as Wordnet. Wordnet is a comprehensive database which follows the standard “dictionary database” structure which is based upon the lemma, the most basic form of a word. It tags the lemma to its other synonyms, lexemes and inflected variations to obtain all the information of the word. Although implementing Wordnet into nTrader project will help the group to save a huge amount of time, the group would still need to modify certain areas of Wordnet before it is usable for the nTrader project.

Besides choosing Wordnet as the database, the group has also researched on various tools for web application development. The group have eventually decided to use PHP as the main programming language in the project due to its high compatibility with the database and its powerful web development framework that allow the group to easily create web pages. All created web pages and applications will then be hosted on a cloud server that is provided by Amazon web services (AWS). The group planned to use AWS Elastic Compute Cloud (EC2) technology as the main web host for the nTrader project. The reason for doing so is mainly due to EC2 high scalability and management capabilities which allow the group to better manage and handle resources.

## Analysis on choice of News Provider

According to the full Literature Review, out of the three; Thomson Reuters, The Wall Street Journal and Bloomberg, we had chosen Bloomberg as our source for financial news due to its reliability and preferred by user.

After weeks of implementation, our nTrader web application system had successfully analyse Bloomberg’s news with no difficulties. Since Bloomberg is currently our only source, hence in average, this application captures approximately 10 news containing the word, ‘stock’, per day. And with our continuous monitoring, we realized the news captured are of no repeating and concentrate on regional stocks, for instance, “European Stocks Drop From 22 Month-High; Richemont Falls” and “U.S. Stocks Rise as China Optimism Tempers Tech Slump”. Hence, this definitely increases our accuracy in calculating Country Sentiment Ratio and plotting 30 Days News Overlay. And we also saw that Bloomberg’s news incredibly matches with our algorithm with little flaws, where expected Signals and Summary are generated with few errors. Thus, we are confident that our results are highly trustworthy and maybe even use by user for references during decision making.

However, we also believed using Thomson Reuters as our source may also return similar results since both Thomson Reuters and Bloomberg have the same Market Shares in the Market Data & Analysis Industry. Thus, if there is a chance for future development, we will definitely give this a try.

## Algorithm to Derive news sentiment

The fundamental nature of the nTrader is dealing with news analytics and its applications within the investment realm. Therefore the constant upgrading and refining of nTraders’ news analytics algorithm is necessary to keep with the ever changing dynamics of the financial markets. There are many existing research literatures on various algorithms for computing news sentiments which have been tested by professionals and generated very positive results. Therefore to increase the reliability and accuracy of nTrader’s news sentiment algorithms, it is essential that the SOLY team conducts extensive study on these algorithms and either make use of them to form the basis for formulating new proprietary algorithms. Currently the SOLY team has identified four algorithms that they wish to look into in the future: Naïve classifier, vector distance classifier, discriminant-based classifier and the adjective-adverb phrase classifier. All four of these algorithms have very different theoretical foundations for computing news sentiments thus they can be implemented alongside one another without inducing any bias. For more details on the four algorithms, refer to the full literature review on *“News Analytics and its Applications in Finance”*.

# PROBLEM RECOGNITION

Singapore was ranked second in the whole of Asia Pacific in terms of financial in the area of investment knowledge, money management, and financial planning (Property Guru, 2012). However, the younger segment of investors belonging to Generation Y possessed the least financial literacy skills considering their level of education versus adults investors. Survey findings by MoneySENSE (2005) had revealed that younger population of investors, especially those new to investing, are not as financial literate as compared to Generation X adult investors, scoring the third lowest mean literacy score of 62 compare to 71 for adults, refer to Illustration: figure 3 for the statistics.

Financial experience is one that accumulates with age. However, the key limiting factor to Gen Y investors’ financial experience level may be stemming from their over-reluctance in the exposure to financial stock market and preferring only conservative investments e.g. savings account (The Wall Street Journal, 2012). In terms of investment returns, this is an unfortunate development, as Gen Y investors have a longer investing lifecycle which allows them to ride through the waves of volatile stock markets.

Even in investing, younger investors also tend to make common investing mistakes. They are reading financial updates instead of analyzing updates, and are speculating instead of investing and performing careful research (SiliconIndia, 2012). As a result, they are constantly lagging behind market changes.

A brief demographic statistics for this segment of investors explains reason for such behaviours (The Echo Boom, 2010). They are:

* starting out their careers and therefore do not have luxury of time to research on their investments;
* university graduates with average debt of S$26,000 upon graduation and seek financial markets as another avenue to pay off their debt; and
* non-finance university graduates having little or no financial knowledge, thus not having the capability nor capacity to make decisions based on financial news.

This segment of young investors will need a package of tools that can support their investment decision-making process, make news easy to analyse and financial knowledge easy to acquire.

# REQUIREMENTS AND PROPOSED SOLUTION PACKAGING

## Analysis of user requirements

Highlighted in this report problem definition, the stakeholders the group’s nTrader project is targeted at are young investors, mentally or physically. Typically, this includes Gen Y investors’ population and beginner investors. This main segment of our target market has little financial literacy education and does not have the luxury of time nor the capacity to analyze investments and financial news; they often read updates in news about market prices rather than focus on important events that should be analysed. In addition, majority of the investors belonging to Gen Y possessed a fearful attitude towards investments and thus has limited exposure to investment knowledge. Taking their problems into considerations, the group summarises the following list of users’ needs.

* They require a system that analyses news and informs the important events or part of the news, thus allowing them to stay focus on what is important within financial news.
* They want to know immediately whether a market update (news, journal article, press release) is a good news or a bad news without diving into too much technical details.
* They hope to encounter as little technical financial terms as possible when being updated about financial investment.
* Financial literacy of the target segment should be stimulated with the use of the create project solution.
* They require a simplified approach to make use of financial news and updates as part of their trading strategy.
* They require a system that value adds and aggregates data from existing sources and puts these data into perspective for them. Information must be readily accessible for them as they do not have the time nor know where to obtain this information.

## Project solution and feasibility

### Solution design

Based on the identified needs of beginner investors, the group envisioned a package of solutions called “nTrader” which will address beginner investors’ needs as well as meet our project scope, and budget. Project nTrader is envisioned as a web application which will allows beginner investors to appreciate impact of financial news to financial stock markets. The group defines nTrader as having two levels: The “system architecture level” which forms the foundation and design of nTrader, and the “system solution level” which are the functionalities and products offered by nTrader.

In terms of the system architecture level, the group deploys the Amazon EC2 (Elastic Compute Cloud) server to host, develop, and test to-be-developed solution. The group rationale for cloud server deployment versus dedicated private server is the effectiveness in terms of collaboration in web design, development and testing, as well as the flexible cost structure brought about by cloud server which is incurred at a pay-per-use basis (currently running on 1 year term free trial version). Dedicated private server requires costly maintenance and one-off hardware costs and changes to server cannot be implemented directly as it is not easily accessible remotely compared to cloud server.

As for database, the group deploys the free MYSQL database solution using the relational database model. The benefit of this is that database records will be unique and data can be retrieved easily by referencing a unique primary key which improves user experience in terms of data load time. Data inconsistency problems that results from the dependence of third party data can also be solved with the relational database model, as data format for table columns are specified. More importantly, the group can save costs in terms of costly data storage as data in the database grows – the said model is able to reduce the amount of data stored compared to using a flat relational model. In addition, the group’s database follows the logical data modeling approach (Refer to Illustration: figure 4). Through this approach nTrader database will be conceptually categorised into distinct categories with relationships represented between each categories. Our main motivation for using this model is to give the group a high level understanding of nTrader features. Additionally, having this logical design as a foundation will also allows the group to expand their database in a scalable manner mainly though the structure of these distinct categories.

For better navigation experience within nTrader, the group also deploys Alternative PHP Caching (APC) as part of its architecture. APC is a free solution for caching and retrieving cached programming variables. In order to improve user experience associated with webpages loading speed, APC was used. Other components of nTrader architecture include JQuery Programming framework, PHP, and Javascript to create nTrader system functionalities. To deliver the most of dynamic webpages, the group also uses AJAX (Asynchronous Javascript and XML) techniques to communicate with server at the client side. The APC, MYSQL, and Amazon EC2 Server however are the core components of Project nTrader solution architecture. There are also other applications of these components of nTrader architecture which extends over the system architecture level of nTrader to explicitly support system solution level. More will be discuss at the following system solution level.

The system solution level is the level where the needs of beginner investors will be addressed. This level includes web files, interface images (picture files), and APIs (Application Programming Interface) files which interact with each other and sometimes with the system architecture level, in order to deliver to users the final nTrader solution. The following table lists the design functionalities which are the results of these interactions between system solution level and system architecture level, the design rationale, objectives, and methodology of implementation.

|  |  |
| --- | --- |
| S/N | Functionalities |
| 1 | The system will be able to inform the important events highlighted within the financial news. The aim of this function is to focus the attention of beginner investors to key financial events that can move the financial markets instead of fixating at market price fluctuation reported by the financial news. The key objective of this function is to enable beginner investors to cultivate a healthy interest in identifying, understanding and analyzing key events found in financial news.  The success of this function is dependent on a Natural Language Processing (NLP) algorithm that can extract subject financial entity, directional movement of the entity, and the event that causes it from a sentence. From NLP algorithm methodology research, the initiation of the algorithm alone increases the execution time of web files that executes the algorithm. The reason being NLP algorithms have to initialize its own internal dictionary before able to analyze natural language. In order to improve script execution time, the group caches the internal dictionary of the NLP algorithm using the system architecture level APC feature. |
| 2 | The system will be able to provide news feed (current day news or archive) and display news sentiment (bull/bear/neutral) for each of the news displayed base on its headline, and the subject mentioned in the headline. Each of the news should come with an introductory paragraph. The aim of this function is to provide beginner investors with a news sentiment indicator for their reference, and without them having to go in-depth into technical news details to simply find out if the news is of a certain nature. Bloomberg news will be used as data source for the design of this function as their news are standardize and edited by news editors before being published to the Bloomberg official website. The sentiment indicator will also be used to perform BI functions in charting tools which will be explained later on.  In extracting financial news data from news provider(s), the group will use regular expression pattern matching methodology to retrieve news. This will be done by searching news details in websites’ source codes. In generating the sentiment indicator signal the group also uses regular expression pattern matching. The group will take a sample size of about 60 financial news headline and identify verbs commonly use to indicate directional movement of financial instrument. We believe pattern matching of these verbs in news headline will allow us to derive with accuracy the news sentiment. Similarly, the group will also manually identify the common subject covered by the sample headlines and tag subject with financial instrument in our MYSQL database. The matching of news subject will also be done through regular expression methodology. This way, the system will be able to generate a sentiment indicator for financial news for a particular mentioned financial instrument in news headline.  This function will rely on the APC feature to store infrequently change regular expression pattern which will identify the subject of financial news, and the directional verbs that indicate its directional movement in the financial news.  As for archiving news to support ability to fetch older news and perform BI, the group will rely on system architecture level’s server event scheduler – crontab. The group will use “crontab” to schedule the function of news archiving. This function will insert into MYSQL database data about archived news such as news headline, captured sentiment signal and the affected financial instrument, news content, type of news, and news date. The group will use similar methodology that display these data, but instead transform this methodology to inserting the data to our database. Similarly, the our design requirement for archive news is the ability to perform BI function on archived news, as well as allow users to searched for relevant news based on the news content. |
| 3 | The system will identify financial terms when displaying financial news and the identified financial terms will be emphasized with highlighting. On click of the highlighted financial terms, user will be able to view its definition and/or add the word to his account for review at a later time. The aim of this function is to reduce as much as possible the technical complexity of financial terms within financial news such that beginner investors can understand financial news without multiple external references to a financial dictionary such as investopedia.  To create this function, the group will need to extract financial jargon data (term, definition, and source URL) from a financial lexicon website. The group chosen financial times lexicon due to its transparency in reporting its lists of financial jargons. There on, the group will use regular expression matching to identify financial jargons and its positions within the financial news, which when combine with JQuery and CSS programming framework, will allow us to create an effect of a popup definition. |
| 4 | The system will have a one-stop portal that allow users to manage jargons added, browse meaning, and play financial word games available. The design of this portal seeks to improve beginner investors’ financial literacy level by exposing them to interactive games.  The group recommends the design of 2 financial word games which will most utilize the large number of financial jargons data that the group extracts has from in financial times lexicon. They are namely “Find The Jargon” and “Guess The Jargon”.  For the design of the “Find The Jargon” word game, user will be required to identify financial jargons with a randomly generated financial news passage. The purpose of this game is to let beginner investors encounter new financial terms within financial news and thus improve their financial literacy. Also through this game, beginner investors will read financial news as they find jargons in the news in an attempt to complete the game. The objective of this game is thus to encourage reading of financial news and well as improving financial literacy.  “Find The Jargon” uses the list of financial jargons in our database in order to match existence of the each jargon in financial news. The group uses regular expression pattern matching to find jargon matches. As it is estimated that the number of financial jargons can exceed over 10,000, the initialization of the word game will take tremendous database resources and slow down the game initialization process. The group will use the APC feature to cache the entire list of financial jargons.  For the design of the “Guess The Jargon” word game, users will be required to guess the jargon with only the jargon definition as a hint. The jargon that will be guessed will be retrieved randomly from the financial jargon database. The game will adopt a “hangman” style interface, where beginner investors guess the word letter-by-letter. The main aim of this game is to increase their vocabulary of financial terms and the objective of the game is to give beginner investors a strong foundation in terms of financial literacy.  The educational portal will also have a function that allows users to search and browse meaning of words and viewing of the full definition of words will take place outside of the site (financial times lexicon), mainly due to intellectual property considerations. Through this portal beginner investors can also view words that he added under his account or remove words that he added as favourite. |
| 5 | The system will have a stock graph chart tool that overlays related financial news with market prices of financial instrument. The aim of this function design is to allow beginner investors the ability to easily add in the consideration of financial news to their trading strategy. The objective of this function is to encourage beginner investors to take financial news into account when performing trading. This function will also have other common technical indicators such as Relative Strength Index, Stochastic, Simple Moving Average.  For this function, the group will use stock charting APIs from HighStock Charts as HighStock Charts official website provides extensive documentations for its APIs along with gallery examples which the group can consider. The interface scheme also matches with our nTrader solution concept with is a “minimalist and aesthetic” design. |
| 6 | The system will have a function that allows beginner investors to take a macroeconomic approach in investing. For the design of this function, the system will be required to aggregate data from relevant sources that will allow beginner investors to make effective investment decisions. For this function, key elements that should be present are a map chart that allows visualization of macroeconomic areas and economic indicators. The map chart should be able to indicate the sentiment of a country through different color shading, and when a country is selected, the country economic indicators should also be displayed.  For the design of this function starting from the map chart, the group proposes the use of Google Geomap API due to its interactivity at the client side and the extensive API documentation available and easy-to-implement coding. As for data that indicates country sentiment, this will be derived from sentiment indicators that are tagged to archived financial news. Each archived financial news are each tagged to a financial instrument and we will tagged this financial instrument to its respective country. Only financial index instrument that indicates the health of a country will be considered. For example, the index for S&P 500 would be tagged to United States. Positive news about S&P 500 would thus be equivalent to positive news for United States. Our country sentiment is the ratio between the count of positive news and negative news.  In addition, the group also proposes the design of a country sentiment chart to plot sentiment of a country over time. The group will use HighStock stock chart API to plot country sentiment data.  For data about economic indicators, the group will use data source from world bank which is free for download in excel format. The group can insert these data into database for manipulation. |

### Database security concerns and design

Database security can be easily compromised whenever users at the client side are able to send data requests to the database dynamically. The group in terms of programming framework uses the Javascript and AJAX (Asynchronous Javascript And XML) technique to communicate with the database at the client side. As the group design the nTrader project solution for maximum user interactivity, functionality have to be coded in Javascript, a client-sided programming language. This means that users can easily modify algorithms and functions coded in Javascript to indirectly affect the group database. As far as the database is concern, the combination of Javascript with AJAX calls that validates or perform an action with the database can result in potential threat to the database, termed SQL injection. In the context of Project nTrader, SQL injection – a situation where rogue data is inserted into database or if existing data modified, will occur if data that will be inserted into the database comes from a Javascript variable or html input to be inserted into the database. The group uses four methods to prevent or minimize risks of SQL injection: There are 1) Server-side validation, 2) MYSQL routines, 3) account privileges, and 4) data format constraints.

Through server-side validation, the group can minimize erroneous data from being inserted into the database. For example in the nTrader register page, the group validate the length of the password and username fields before inserting them into database. Therefore even if users modify the codes in the client side, the server side will act as an additional layer of protection to filter rogue data.

MYSQL routines however is the most frequent used SQL injection preventive measure use by the group. MYSQL routines constraints the ability of users or hackers in modify SQL by insisting a pre-defined set of parameterized inputs. The ability of users or hackers to intentionally modify SQL queries in the server side will not be possible will MYSQL routines, as pre-defined MYSQL routines distinguishes parameterized data inputs with routine definition statement. The original intent for the SQL query will not be violated. The group mainly implements MYSQL routines for SQL queries that perform insert, delete or update operations to the database.

Through account privileges, the group also strengthens database security. For the purpose of executing SQL select statements, the group had created a nTrader database user account with only select privilege to the database. When viewing data from the database, only the nTrader database user account will be used. This prevents the database attacker from forcibly modify the SQL query at the client side into insert, delete or update operations which affects the database.

The last method used by the group in preventing erroneous data in database is through specifying explicit data format in the database tables or in MYSQL routines parameterized inputs. This prevents data or with erroneous format (e.g. timestamp data with alphabetical characters) inserted into the database or MYSQL routines to be able to execute successfully, protecting the database.

The group also pledged to protect users data especially password information as it is common for users to tend to repeat their password use in other external registered websites or services to setup their nTrader account. The potential disclosure of such sensitive data will cause potential damage to users and their loss of confidence in the group’s future products and services about nTrader. The group protects password data by encrypting them using the SHA (Secure Hash Algorithm) 512 password encryption algorithm. The group used the SHA512 encryption algorithm as it is internationally recognized as a U.S. government standard by the Federal Information Processing Standards (FIPS). It is also the most secured encryption algorithm currently versus the likes of MD5 encryption as it is impossible for the former encryption algorithm to create collision vulnerabilities, a situation where two different password inputs will enable the attacker to log into the user account. SHA512 encryption is irreversible which also reduce the risk of password decryption to zero.

# FEASIBILITY STUDY

## Technical feasibility

The group determines technical feasibility base on accessibility and utilization rate of these tools and resources. There are three major technological components that made up the nTrader:

* Computer languages
* Web server
* Database

In addition, the group initially also wished to incorporate natural language processing algorithms into nTrader. Currently the biggest constraint that the group has with regards to this matter is learning curve. As natural language processing is a foreign subject to the group, there is substantial amount of risk taken if the group were to gear the nTrader platform towards it. Furthermore, conventional practices of setting up finance related NLP databases require the aid of financial linguist. The group however is confident of producing an amateur natural language processor within nTrader due to their competent nature in the field of technology. To accommodate for future developments, the group will explore and set up the foundation layers of infrastructure essential in the implementation of powerful natural language processing.

Computer languages feasibility

Since nTrader is a web-based application, to best develop nTrader project in terms of usability, processing efficiency and effectiveness, and to meet general requirements that constitutes a website, computer language required are HTML, CSS, PHP, Javascript, SQL, and Python.

As project nTrader will require an element of NLP algorithms, Python language is the most suitable choice to perform such algorithm; python has the supported libraries to support efficient text parsing and reduce development time. Based on the group’s technical expertise however, Python programming language is not possible as the group has no previous exposure/knowledge to the language syntax and learning time would exceed project time frame. The group performs a workaround using PHP programming language for NLP algorithms. The group had previous exposure to HTML, CSS, PHP, SQL and Javascript computer languages.

Web server

The Amazon EC2 server was chosen to enable Project nTrader as a remotely accessible web application. The following Amazon EC2 server specifications make development of Project nTrader feasible (Amazon, 2012).

|  |  |
| --- | --- |
| Specifications | Comments |
| 750 hours of Amazon EC2 Linux† and Windows Micro Instance usage (613 MB of memory and 32-bit and 64-bit platform support) | Enough hours for server to run continuously each month, allowing nTrader to test its applications on different operating system thus enhancing the reliability of nTrader. |
| 750 hours of an Elastic Load Balancer plus 15 GB data processing. | This ensures nTrader will not face any downtime caused by instance failures as the elastic load balancer is able to detect faulty instances and automatically reroute traffic to non-faulty instances until the faulty instances are fixed. |
| 30 GB of Amazon Elastic Block Storage, plus 2 million I/Os and 1 GB of snapshot storage. | This storage device exists independently from the life of an instance, thus preventing data loss due to any failure from the instance. |

Database

As nTrader requires huge amount of data to carry out its operations, it is important that we assess the possibility of obtaining this large set of data. The primary data that nTrader needs are: financial news articles, market price data, economic data and financial lexicons. The group has identified the main sources of these data and has devised ways to extract them.

The three main sources of data are the following:

* Bloomberg – Where we extract data for market prices and financial news
* Financial times – Where we extract financial lexicons
* World Bank – Where we extract economic data

Although we know where to obtain the data, however nTrader has a time constraint of 10 weeks and to extract huge amounts of data manually under this constraint would be infeasible. The group did a calculation and the amount of data that we needed to make nTrader operational ready for just the feature nTranslator would be an approximate of 12000 rows of financial terms and jargons. The group is able to construct algorithms using excel VBA to automate the process of data extraction from these websites. As data are located on different webpages, our algorithm is able to browse through every webpage and extract required data based on pre-defined specifications (refer to Illustration: figure 5 for a screenshot of the data extracted). Based on our estimates, this would reduce time taken for data extraction to a minimum of 8 hours and a maximum of 12 hours.

## Economic feasibility

The main costs for project nTrader would be the maintenance of its cloud server as Amazon’s EC2 cloud server is only free of charge for a limited time period of one year. Since Project nTrader was initiated under a zero-cost budget constraint and is not legally allowed to engage in commercial activities due to legal requirements arising from the free usage of its third party data, possible future sources of funding to sustain the website will be limited to the subscription fee from users or advertisements.

As compared to a dedicated server, the Amazon EC2 cloud server has a relatively better costs structure. The cloud server charges only for resources that the user uses whereas a dedicated server has fixed pricing structure. Although cost rates are normally higher for cloud servers on per hour basis for the same amount of resource usage, nTrader only require very minimal amount of system resource to be fully operational (refer to literature review *“Technology”*).

The cost for nTrader on the EC2 cloud server is as follows (refer to Illustration: figure 6 in Illustration):

* Micro-on demand instance on windows operating system= $0.020/hour
* Rates for data transfer are as follows: $0.11 / 1 mil Input/output request

If we aggregate the cost to per month basis, on EC2 cloud server, nTrader would costs a total of $14.67 per month. As compared to a dedicated server which nTrader is able to operate on, ServerPronto, the cheapest existing server available is at $50 per month. Therefore for nTrader, the most cost effective solution for web server would be the EC2 cloud server.

## Legal feasibility

As nTrader uses third party data, it is important that we obtain the legal rights from these parties. Since nTrader uses a web crawling method to retrieve data from third party websites, one way to check whether the third party sites approves or disapproves web crawlers of accessing their contents is to make used of the Robots.txt. The Robots.txt is a file which gives instructions about their site to web robots (e.g. web crawlers). These instructions will indicate the content within the website which the web owner does not wish to be used by these web robots. Based on Robots.txt, Bloomberg allows web robots to use its news contents and market price data (Refer to Illustration: figure 7). Financial times as well did not disallow the extraction of its financial lexicons and definitions (Refer to Illustration: figure 8).

As for World Bank, they explicitly state on their terms and conditions that their data are offered free to the general public.

To have a safer confirmation of the legal rights to use Bloomberg’s contents, the group contacted the legal department of Bloomberg for permission of use. At the point of writing this report, the permission request form that has been submitted is pending approval (Refer to Illustration: figure 9).

## Operational feasibility

Young investors do not have the luxury of time to conduct investment research. Therefore, nTrader addresses this problem with the news sentiment indicator concept where sentiment ratings will be generated base on news headlines. Reader will be able to grasp informational context of the news without spending time reading the full content of the news.

Investors that are unable to effectively visualize the relationship between news and financial market prices also make nTrader the desired solution for such problems. For example, nTrader has a Business Intelligent graph and map component that gives users the visualization to link quantitative aspect of news data with financial market prices.

In addition, nTrader will be provided as an open-source web application for retail-level investors. In the current market, there is a shortage of free or affordable financial news analytics application. The absence of competition in the free or affordable category of the financial news analytics application market with allows project nTrader to be more marketable.

# PROJECT TRIALS & TESTING

## Difficulties faced

### Graphical User Interface

One major issue that the group encountered at the start of the project is the amount of control the group has over the server. As discussed in earlier section, nTrader uses a cloud server hosted by Amazon Web Services (AWS). Since it is the first time the group members have hands-on experience with a cloud server, members were unsure on how to actually use it to host web applications. This problem was easily overcome as the group slowly picked up the Secured Shell (SSH) skill that is used to communicate with the cloud server. AWS provides a SSH console that serves as a communication tool to the server. The group uses it as the main communication device to control the server to perform system task such as installation, updating and internal file management.

However the group soon realize that the only way of using the server was also to use the SSH console which does not provide a proper system interface or graphical user interface. This became an even greater problem when the group wants to perform database management where data are not displayed through a proper GUI and various complicated commands have to be used before SQL statements can actually be sent (Refer to Illustration: figure 10 for image of SSH image). File transferring was also a difficult task as it can only be done through the SSH console which requires the group to type the full path or slowly navigate through the folders before being able to select the specific file to transfer (refer to Illustration: figure 11 for screenshot of SSH file transfer console).

After several research and discussion, the group decided to implement a database management interface known as phpmyadmin and have each group member install a third party file transfer protocol, FileZilla , which can serve as alternative communication tools to interact with the web server. The implementation of FileZilla was simple however unlike most servers, AWS Elastic Compute Cloud (EC2) uses a different file transfer method which does not use password to identify user and instead uses a key that is created by the group. This provides a better security however this means that the key has to be shared among all the group members otherwise the other group members would not have access to the server. The group uses a shared folder service that is provided by Dropbox and through this, the group was able to efficiently share the access key and documents that are used in the project.

The installation of myphpadmin was slightly complicated as the server does not recognize third party software that is not pre-loaded into the AWS image used when creating the cloud server. In order to install phpmyadmin, the group had to use SSH to change certain folders and file ownership and use the FileZilla Client to send the installation package to the server. The group also had to setup a MYSQL database on the server before phpmyadmin was able to work. After installing new GUI on the server, the group was able to better manage data on the database and the new GUI serves as a new platform to send SQL queries and view table structures.

### Data gathering

During the initial phase of the nTrader project, the group has identified that in order for nTrader project to proceed smoothly, we would need a large amount of financial, text and news data. Therefore, the group has initially planned to make data collection a top priority and allocated a substantial amount of time dedicated to finding financial terms, definitions and compiling all the gathered information into a proper database with a well-defined structure (Refer to Illustration: figure 12 for database structure). The process of finding large amount of relevant data that can be used was difficult and it is even harder to check if the data gathered is reliable.

The group originally thought of using manual identification method to identify financial jargons manually and look up on the internet for each identified jargon’s definition. By doing so, the group can ensure that the financial terms found would be highly relevant however the data found would not be credible since there are multiple sources and it would be difficult to check for data source validity. Furthermore the members in nTrader project are not professional linguist or financial experts thus words that are identified by the group are also not credible. Eventually, the group decides to look for existing financial lexicons and collect the data from reputable financial lexicons.

This created a new problem of efficiency and resource management for the group. As there are large amounts of data existing in financial lexicons, manually copying data over to the database would not only waste large amounts of manpower but also become mentally taxing to the group members and inefficient. In order to avoid this situation, the group has thus research on the possibility to automate the process by creating a web crawler. A web crawler is able to browse the web pages of the financial lexicon website rapidly and by reading the source code of each page, the group managed to identify similar patterns and created a web crawler application that can extract financial keywords and its related definition (Refer to Illustration: figure 13 for web crawler application VBA).

Although the web crawler application has helped substantially by reducing the amount of time taken to gather data, it is not a perfect solution for data collection. Most of the data collected by the web crawler are usable however there are some data that have flaws such as missing definitions, irrelevant financial keywords and incorrect definitions. Another major problem that the group found was that the web crawler overlooked the word relationship between financial keywords as some financial keywords may have similar meaning. This sparked another discussion among the group members and after several debates, the group decided to further refine the web crawler and extract financial keyword relationship and compile them to a second database table. A second application was also developed to clean up the current collected data while a group member was assigned to verify each collected financial keyword and remove useless keywords.

Creating the web crawler has helped solve the problem of data gathering however all the data gathered by the web crawler are not inserted directly into the nTrader database and instead they are stored on excel sheets. Data has to be stored on the database in order for it to be used and inserting data manually into the database will be as difficult as gathering data. Therefore the project group has decided to automate the process. The project team initially wanted to connect to the database directly and insert data however that was not possible as the database was locked up in the EC2 server and none of the group member know how to access it. However SQL insertion can still be done through phpmyadmin interface and the group thus created a third application which is able to convert multiple rows of data in excel document into SQL insert statements (Refer to 8 for SQL query creator). With the help of the SQL generator, we would be able to insert large amount of data in batches by sending batches of SQL statements.

Through all these applications that the project team created, data gathering became much faster and easier. The group was also able to collect 12,000 financial keywords in a small amount of time which would not have been possible without these side applications. With a database that contains such a large amount of keywords, the project team also opened up many new possibilities and features such as business intelligence and word learner portal service that would not be feasible without large amount of data.

### Change Control Management

As the project progresses into the mid-execution phase, a new problem surfaced which caused minor conflicts between group members. This problem was caused by the lack of a proper change control management framework within the group. Change control seemed to be a low concern for the group initially as decisions were made as a group and once decisions were finalized, actions were taken immediately and no further changes would be made. However when development of the website began, each of the group member was given their own task. Initially, after the completion of individual task, each member will submit their work to the shared folder provided by Dropbox. The key developers would then take the work and compile them before updating the website.

Although all members of the group had server administration power, most of the group member only used the power for extraction of files to their own internal development environment. The group was afraid that by using the administration power to rewrite files, important data may be rewritten and work may be lost. Thus the group came to a general consensus that only the key developers will execute the file rewriting process to update the web site so that when a work loss scenario happens, the key developers can better understand what was rewritten and reverse the process.

The initial implementation of these rules seems to be effective. However, as each group member has their own preferred working hours, a daily one-time update of the server was difficult and instead the server was updated on an on-demand basis. Furthermore, synchronized updating of work was impossible since only the key developers were the ones updating the server and that created version control problems for the web designer and other members. Since the web designer is unaware of the latest changes made to the website, his version of the website which he made changes on was outdated and when it is submitted to the key developer for compilation, the website got rolled back to a previous version. This created more work for the key developers as the key developers have to compare the changes between the two versions and recompile them into a new version and over time this proved to be an inefficient change control framework (Refer to Illustration: figure 14 for initial framework).

The group became aware of the issue with the change control framework and started to make changes to respond to these issues. Firstly, the group decided to create individuality between files and thus only one group member can work on the file at any point of time. In order to implement this, the group decided to improve communication between member and set up a work flow log to keep track on the current and future task that each member have to do. Changes also have to be updated real time and each member was allowed to execute the rewriting power to submit files to the server when they completed their work. To prevent the work loss situation from happening, group members were required to make a copy of the file they would be changing before rewriting it so they would be able to roll-back the file to its previous state if something goes wrong. Finally, group members were required to refer to the server version of the file before they start working on it. This was to ensure that the version that was being worked on would always be the latest version (Refer to Illustration: figure 15 for updated change control).

The effect of the implementation of the new framework was not immediate as some group members were unaware of the new policy however as the group became more used to the new framework, the number of issues relating to poor change control reduced and work efficiency within the group improved greatly and group members would not need to obtain updates from key developers before working on their changes.

# EVALUATION & USER FEEDBACK

## Evaluation of project processes

Many decisions were made in the project discovery phase. Our original plan was to create a portfolio and trading analytics, however the competition in the field of trading platform are intense which causes the group to deviate from the norm and start on the nTrader project. Knowledge and skills gained has allowed us to create ‘text-parsing’ algorithm that analyses news and extracts important information. While Natural Language Processing algorithm was part of the project plan, NLP algorithm was not implemented fully. However it was a learning experience in understanding how texts can be translated automatically according to context was valuable.

Next is planning phase. After deciding on the nTrader Project which started off with a news analytics concept, roles are assigned to each members; logic programer, web designer, Database Maintainer and Project Manager. Subsequently, each role is assigned with task and this is where Project Manager steps in. He is in charge of ensuring others are following the schedule (Gantt Chart) and the deliverables are of good standards. However, the major problem in this phase was brainstorming the programming language for our nTrader. Though we have some fundamental knowledge in languages such as HTML and PHP, it wasn’t enough. Thus we planned to learn Javascript, AJAX and JQuery. Furthermore, there was a need for us to brush up our PHP.

The second biggest issue is the development phase where creation of the infrastructure is what our group is proud off. We adopted Amazon EC2 as our host server, Logical Data Modeling database approach as well as database securities such as database routines and SHA512 encryptions algorithms for our database which are of industry standard. Though we faced any difficulties in using EC2 as it is new to us, the group came together and discuss about the solution. Moreover, we have also put in effort especially in the extraction of data. We also created VBA-program that does automatic extraction and insertion of data, decreasing amount of effort required to build up the database.

Moreover in the implementation of our nTrader's algorithms, the group faced a major issue regarding the reliability and accuracy. Thus, the group had again, came together and research on industry standard algorithms that is often used in text analysis. Though there isn't a fixed or suitable solution in our case, we had made refinements such as getting more samples to strengthen the bull-bear sentiment generator.

In conclusion, as we progressed every week, more and more ideas and improvements surfaced. Our continuous improvements and upgrading will leads to success of nTrader.

## User evaluation

The results below are gathered from 16 respondents in a survey (Refer to Illustration: figure 16 for survey questions). Charts for each of the following questions are available in Illustration: Figure 17 to 21.

|  |  |  |
| --- | --- | --- |
| **Question** | **Financial Savvy** | **Non-Financial Savvy** |
| Are the bull/bear indicator found in the homepage useful? | 3.5 | 3.0 |
| Do you find the charting tools in nIntel useful? | 3.2 | 3.0 |
| Do you find WordLearn HQ portal conducive in learning financial terms, especially the word game? | 3.4 | *2.3* |
| Do you find nTranslator feature helpful in understanding technical terms within financial news? | 3.7 | 2.7 |

\* Underline – Indicate lowest rating among all applications

\* Highlight – Indicate highest rating among all applications

In terms of the helpfulness of each application, for Financial Savvy users, we see nTranslate having highest scoring. This means that financial savvy users are keen in expanding their financial knowledge and that the application was useful. On the other hand, nIntel has the lowest scoring as the group’s nIntel graph feature has it has little functionalities versus graphs from Bloomberg and Reuters. As for Non-Financial Savvy respondents, we see that both nIndicate and nIntel have the highest scoring while WordLearn HQ has the lowest scoring. We previously expected nTranslate and WordLearn HQ as an educating applications specially mean for Non-Financial Savvy. Therefore, due to unexpected result, further clarification will be done in next round of feedback where users evaluate which application(s) is/are the most helpful.

From the graph above, we see that amongst all applications, majority of financial savvy users (43.8%) feel that nIndicator is most help feature whereas majority of non-financial savvy users (42.9%) feel that nIntel is the most useful. These results are different from our expectations as we expected nTranslatate and WordLearn HQ as an educating applications specially mean for Non-Financial Savvy, and nIndicator and nIntel for the more financial savvy since it requires some level of financial knowledge in order to utilise. We conclude that financial savvy users find nIndicator as a convenient tool in indicating whether particular news is Bull or Bear as well as the affected instruments. With this signal, we believe financial savvy users may leverage on these analyses and include as part of decision making process. On the other hand, results from non-financial savvy users are out of our expectation as the chosen nIntel has no educational purposes and is purely meant for analytical purposes. Based on this phenomenon, we conclude that majority chose nIntel due to its interactivity, user-friendliness, and the ability of the graph to relate news with market prices which eliminated the problem of having to read financial news for these users. Coincidently, these are key components that non-financial savvy users look into when using financial products.

## Self-evaluation of Project nTrader

### Strengths

|  |  |
| --- | --- |
| **Comments** | **Respondents** |
| - Layout and Presentation of the Webpage was Clean, Clear, and Concise.  - The Games is entertaining and great visual effects of the Hangman helps to keep users engaged for at least a minute.  - WordLearn HQ section is beneficial as it helps reader to sieve out key words in an article and at the same time educating them with new financial terms. | Low Kai Keng  (FBI gradurate) |

Like what Steve Jobs had said ‘Design is not just what it looks like and feels like. Design is how it works’, and we believe our users feel the same way with our Clean, Clear, Concise layout since we had especially assign a web designer role to a member which focus on design. Furthermore, we understand that the way to attracts users is dynamic and interactive webpage content. Hence, with limited skills set, we made used of interesting images to catch users’ attention. Users will feel entertained and learn new financial jargons at the same time. Moreover, we have also seen the phenomena that when users reading online news, he/she may wishes to drop down some keywords for future referencing. Looking into this, our WordLearn HQ is the desired application which allows users to bookmark jargons to their respective word list and reference it in the future.

### Weaknesses

|  |  |
| --- | --- |
| **Comments** | **Respondents** |
| - Games are limited for time being. Await for the "Coming Soon" games  - nIntel purpose was not clearly presented in the page. Perhaps could include personal sentiment of whether should investors invest in the country or not when they clicked on a specific zone.  - Could have more e-Learnings materials to help beginners on basic investment knowledge for instance how to analyse and read the graphs and different objectives of the indicators. | Low Kai Keng (FBI gradurate) |
| The page layout is not very appealing to users. | Joyce  (FBI) |

nTrader has to build on the knowledge of the above aforementioned comments in order to improve usability. In terms of our applications, an improvement will definitely be stating down instructions to give users clearer idea on how the application can be used as technical indicators such as our ‘30 Days Simple Moving Average’ and ‘14 Days RSI’ are not easily understandable and usable by beginner investors. A help page containing instructions of the used of the website features should be made available. Our website should also be more appealing in terms of its design and flow. Therefore, the web designer will need to have excellent knowledge in the study of color contrasts and web elements placement.

As for WordLearn HQ, due to time constraints, only two games were launched, which are ‘Find the Jargons’ and ‘Guess the Jargon’. If the two games were to lose its replay value, the WordLearn HQ portal will no longer serve its purpose as a conducive learning portal. Existing games should be constantly updated and new games should be launched to retain these users.

# RECOMMENDATIONS

## Technical improvements

### Security

One key area that the group could improve on was security of the data in database as there are still areas where improvement made can greatly match industry standards of security levels. As discussed earlier, the group used the SHA512 password algorithm to encrypt confidential data in the database. Although it is a one-way encryption algorithm, it is still possible for experienced hackers to decrypt the data. The better solution for the group would be to create and use a self-created algorithm which would however requires high level of technical expertise and time which the group presently does not have and thus it was not implemented.

Another area that the group could have improved on was the use of server sided double-validation. Most of the codes used in the current nTrader project relies on client sided Javascript variables to store temporary data. This allow users that have technical expertise to edit the variables stored to falsify data sent to the database, break validation criteria or even perform SQL injection attack to steal data. Since most data used in the current nTrader project is open to public, the group did not attempt to further address these security loopholes however this is a major area that the group could improve on. The group only implemented server-sided validation in registration of users however.

### Cloud server management

Amazon Web Services (AWS) is one of the world’s leading cloud service provider and that is one of the main reason why the group used it for the nTrader project (TechTarget, 2012). Through AWS, the group was able to perform many tasks that would not be possible on other servers. However the group has still not unlocked the true potential and fully leverage on the capabilities that AWS possessed. Currently the group only uses the scheduling, setup and backup features that AWS provide but in the future, the group seeks to implement automatic resource scaling, automated backup and scheduled reporting which will truly display the advantages of using AWS as a cloud server provider.

### Natural Language Processing (NLP)

The group originally planned to implement NLP into the nTrader project to create an accurate and dynamic text reading system which is capable of summarizing large amount of text and retrieving important sentence. This was not possible during the execution of nTrader project as the group found that the system capabilities were not powerful enough and that the WordNet database was incapable on differentiating financial keywords from normal words. Although the group made an alternative database which was able to identify financial keywords, the system could not identify inflected forms of financial keywords due to the lack of inflected form tagging.

In order for the full NLP features to be implemented on the nTrader project, a great amount of time would be required to create a lexicon database that is geared for financial usage. Thus an area that the group could have improved on would be to improve or re-develop a financial lexicon database that is capable of identifying not only financial keywords but also its synonyms and inflected variation. With such a database, the group would then be able to apply more advanced algorithm to provide users more quantitative analysis and values when analyzing news.

### Browser compatibility

Being a web-based application, nTrader aims to use the power of the internet to distribute its content to a large amount of users worldwide rapidly. However due to the large amount of jquery usage in the coding, users that are using older browser such as Internet Explorer would be unable to access nTrader content. This limits the options for web browsers as user that does not have up-to-date web browsers such as Google Chrome or Firefox would be unable to use nTrader. The group should not limit users options and should had used javascript instead of jquery during web development to allow Internet Explorer users to view and use nTrader.

### Screen resolution

nTrader was developed on a widescreen environment (16:9 aspect ratio) as most computers that are available in the current market comes with a widescreen display. Although screen resolution would be of no issue to these widescreen users, users that use older model of monitors that use standard display (4:3 aspect ratio) would not be able to view certain web pages as intended. Due to the lack of time, the project team decided to provide only widescreen support for nTrader. With sufficient time, the project group could perform modification and intensive testing on the code to allow nTrader to support standard screen resolution in the future.

## Non-Technical Improvements

### Gaming

Besides technical improvements, nTrader has many non-technical areas which can be improved. One major improvement that can be implemented in the future would be the gaming concept. In the recent years, gaming has grown greatly in popularity with more and more websites such as Facebook using games as a means to attract and retain users. According to Facebook statistics, the main reason why users play games is due to competitive spirit rather than fun (Foxx, n.d.).

By introducing an incentive system built on top of the gaming concept, nTrader would be able to also attract users and provide a higher replay value for the nLearner games. The incentive system could provide advantages within nConsole to users who are able to demonstrate higher level of financial literacy and by doing so, spark a competitive spirit between users to gain a higher financial literacy level. This will help contribute to nLearner’s purpose and goal greatly by creating a competitive yet fun learning environment for users.

### Personalization

Personalization could be another feature that the project team could better work on. Every user of nTrader is unique and each user has their preferences however the current nTrader is not capable on creating a unique environment for each user such that each user will have a unique identity with features geared towards their preference and personalities. To improve on this area, the group could create personalization features such as settings of the type of news the user would like to view as well as a customizable and dynamic nConsole that the user could design and edit.

### Text-based support

The current nIndicate relies heavily on the news headline format to perform NLP processes on. The group hopes to improve this in the future by implementing a full text-based support for nIndicate such that users could pluck and place any text into the nIndicate processing engine and the engine would be able to analyze and output the financial information that are derived from the engine. By having such a processing engine, the group would be able to provide news sources to user and create highly flexible and reliable nIndicate system. Due to limited project scope, the group is unable to create engine powerful enough to perform this task.

### Multi-Language support

nTrader was set up on a web environment to provide a convenient platform to distribute content and services rapidly to services worldwide. There are various users worldwide however nTrader limits the type of user by only providing content in English. This will thus only allow English literate users to use nTrader which is not the original intention of the project group. Implementing a multi-language support feature is currently not possible for the group as financial data of other languages have to be collected and a new NLP algorithm have to be created in order for nTrader to work. However with a longer time scope, the project group would be able to implement multi-language support which will remove language barriers that restrict users from using the nTrader services.

### SMS support

One of nTrader’s key target audiences are busy beginner investors that do not have luxury of time to read news. nTrader have helped this target audience group greatly by providing them summarizing functions and indicator features to quickly understand news and gain insights. Although these solutions are highly useful, the group has overlooked the possibility that these busy investors do not have internet access on-the-fly which limits their accessibility to these features. To manage this problem, the group should implement a SMS feature so that users would be able obtain insight from the nTrader services while internet access is unavailable.

### WordLearn HQ Improvement

WordLearn HQ is a platform which aims to help users learn words through interactive games and developing individual user personal word database. Since its launch, WordLearn HQ has proved to be an effective educational tool however there are still many room for improvement to enhance user’s learning experience. Although the group has currently planned to launch more games in the future to add variety in the learning process, areas of WordLearn HQ such as the financial lexicon still requires continuous improvement to improve data quality. The personal word database provided in WordLearn HQ also has little functionalities to aid user learning thus the group could add more features to improve the content quality of WordLearn HQ.

### Trading

As the name of the project suggest, the original goal of nTrader is to use news to facilitate trading decisions. However due to the problems and scope limitations the group face, the group was unable to create a highly powerful and reliable processing engine to provide trading signals that users could trust. Without the capabilities of such an engine, the group decided that the trading functionality should be disabled and this removes a great feature of nTrader. However the group is confident on developing a highly capable engine in the future and could bring in the trading functionalities that were intended for this project. With the trading functionalities, users would then be able to use financial signals generated from the engine to perform trading of financial instruments.

### Forums

The greatest benefit of nTrader is its ability to generate positivity signals by analyzing news. Although this is a wonderful feature to have for nTrader, the group feels that nTrader should not just provide a one-sided analysis of a piece of news and instead understand the different perspective of various investors and generate a common sentiment. In order to provide such a sentiment, the group should provide a forum where nTrader users could engage in discussions and understand how other investors think.

The project group believes that learning the application of financial news to financial market is as important as learning the financial theory and definitions of keywords. The forum would serve the purpose of gathering investors and sharing investment thought process which will also contribute to the goal of nTrader which is to aid investor understanding of financial news and markets.

# CONCLUSIONS

In terms of a retail-level investment console framework that is especially targeted at beginner investors, the field of news analytics in order to bring insights to their investments is relatively unexplored. Through project nTrader, the project team opened up this unexplored area and demonstrated the importance of financial news to help investors make better investment decision. nTrader comprises of four main component which are nIndicate, nTranslate, nIntel and nLearner. These components have also brought about great benefits such as:

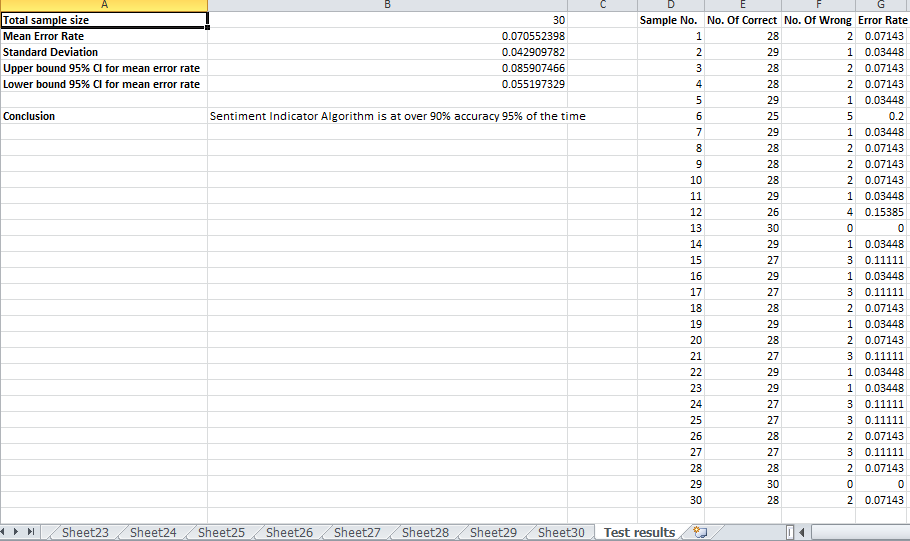
* Improve user’s financial literacy
* Reduce time spent reading news
* Understand financial news better
* Better appreciation on the impact of news and financial markets

All of these benefits were only made possible with the development of the natural language processing engine. With the help of the engine, investors would be able to obtain trading signals and financial insights. The engine also allowed the project group to gradually build up a massive structured database which supports all the operations of the web application. In a nutshell, nTrader has proved to be a highly effective tool to help investors and encourage them to learn more financial knowledge but there are still many areas that nTrader could still improve on. The project team hopes to implement personalization and engine improvement features in the future to mold nTrader into the ideal platform for all investors to analysis on.

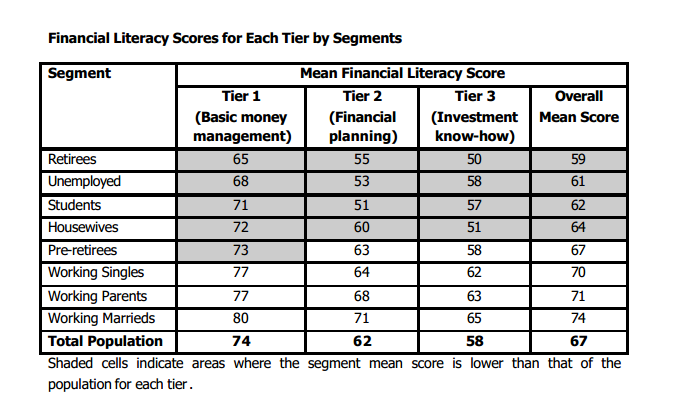
# ILLUSTRATIONS



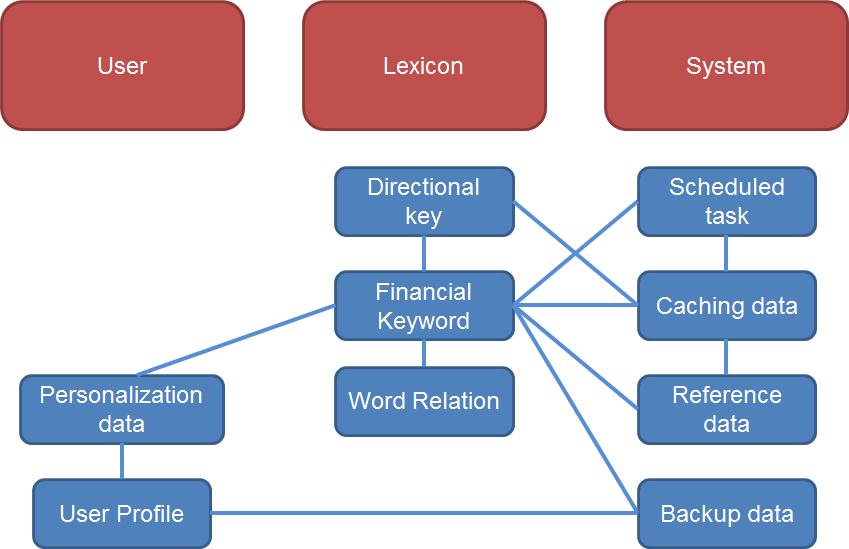
**Figure 1: Simple random sampling of 30 news headlines with the word ‘stocks’ in it**



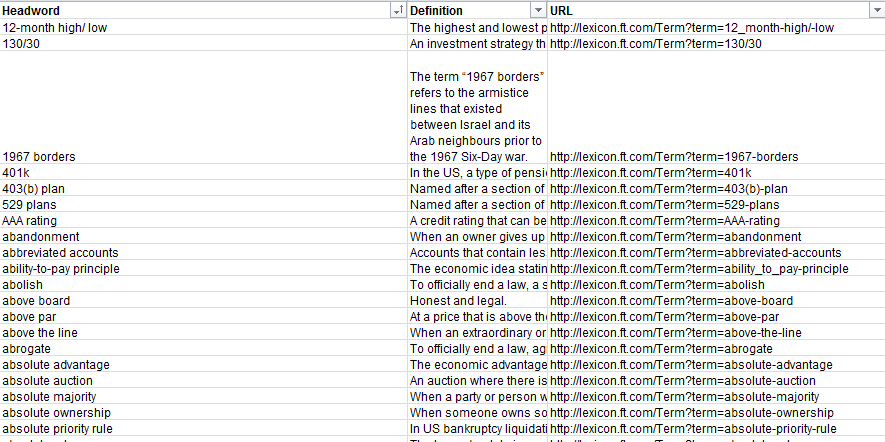
**Figure 2: Accuracy statistic of bull-bear sentiment generator algorithm**



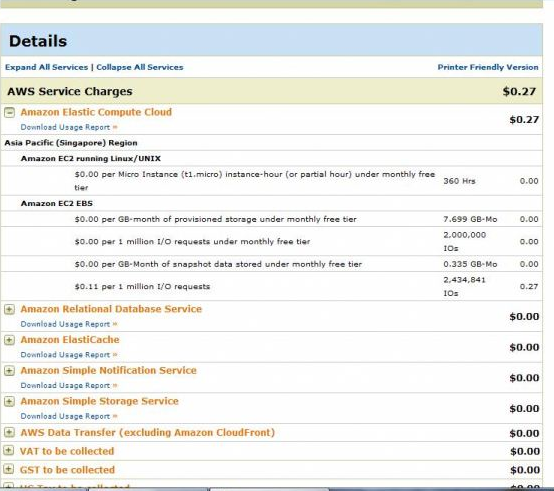
**Figure 3: Literacy scores for tier segments**



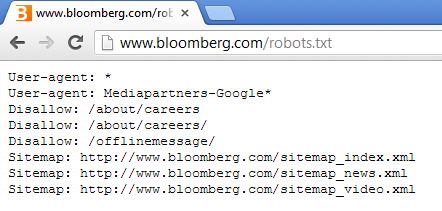
**Figure 4: Group’s database logical data model**

****

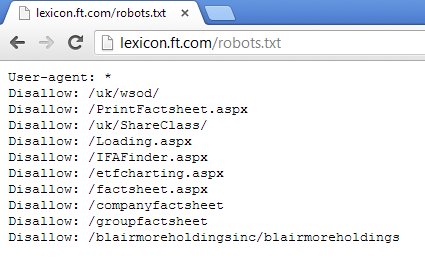
**Figure 5: Extracted data using Excel VBA data extraction algorithm**

****

**Figure 6: Cost of Amazon EC2 Cloud Server**



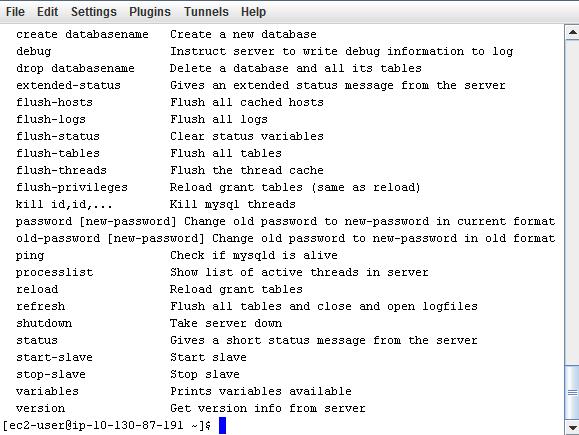
**Figure 7: Bloomberg’s robots.txt**

****

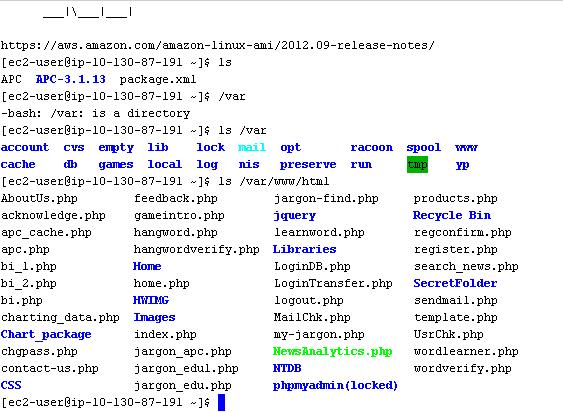
**Figure 8: Financial Times Lexicon’s robots.txt**



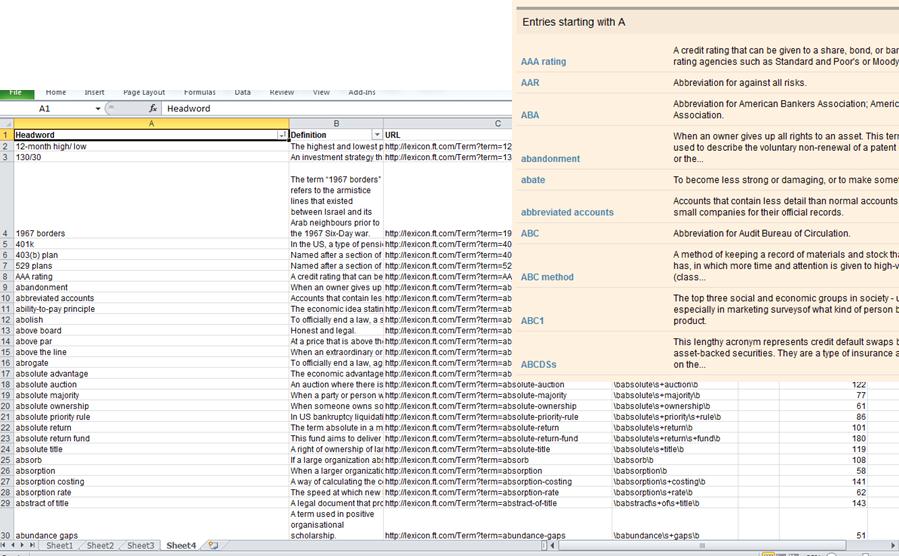
**Figure 9: Bloomberg’s permission request form for use of site materials**



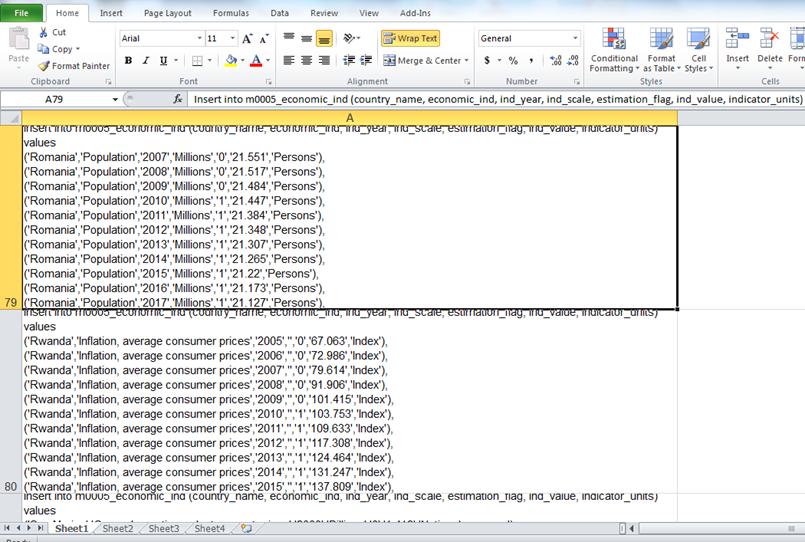
**Figure 10: Secured Shell console image**



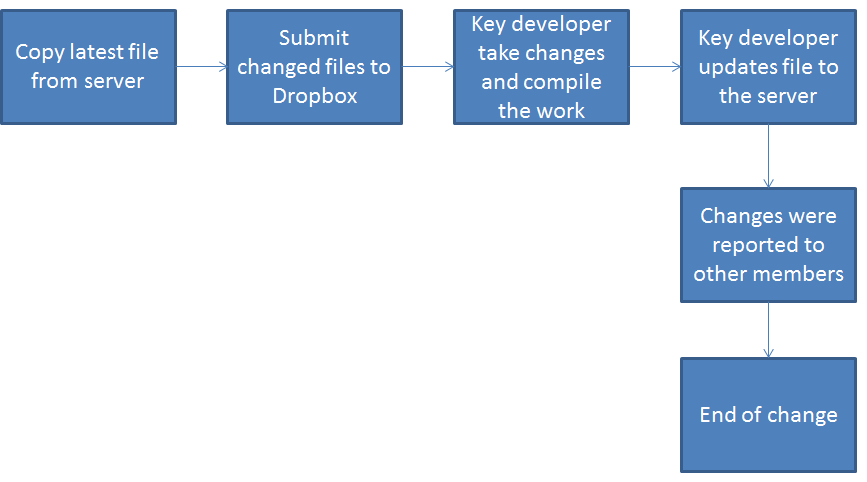
**Figure 11: Secured Shell File Transfer Console**



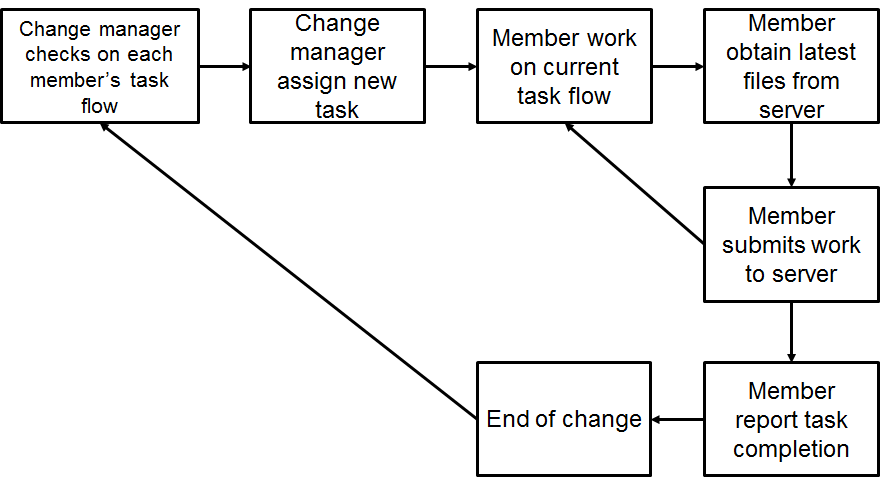
**Figure 12: Excel side application web crawler**



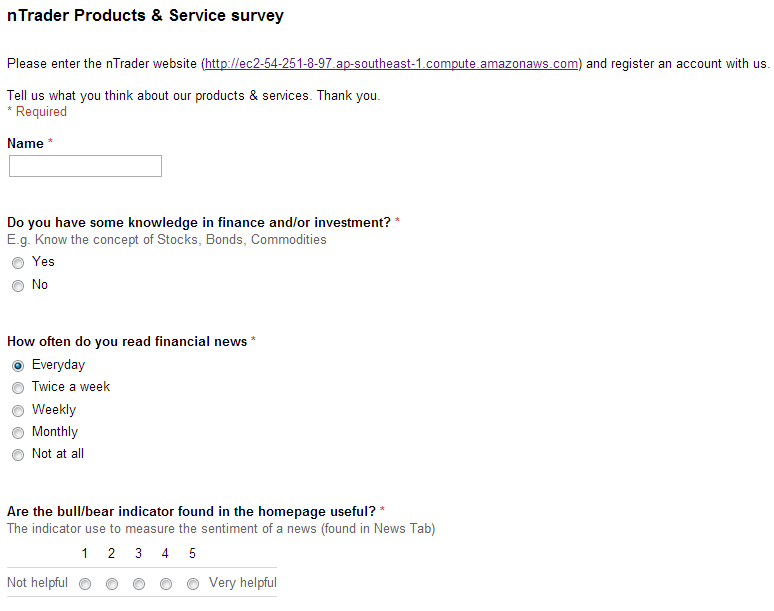
**Figure 13: SQL insert statement batch generator**

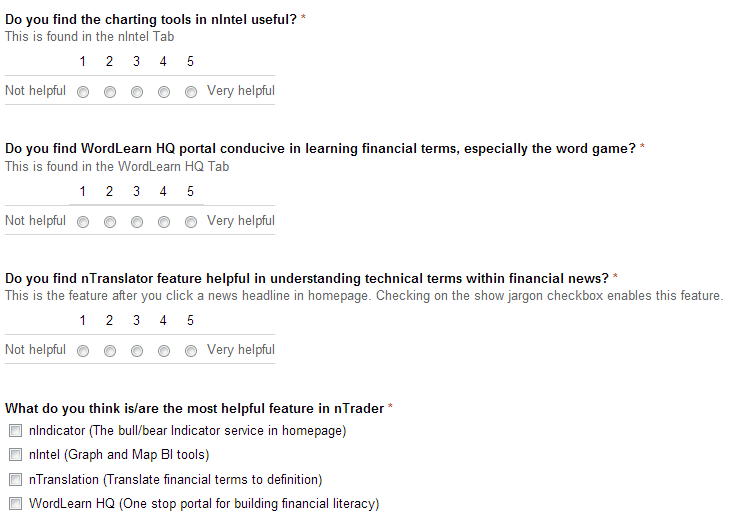


**Figure 14: Initial Change Control framework**



**Figure 15: Updated Change Control framework**

****

****

**Figure 16: nTrader products & services survey**

**Figure 17: Usefulness of nIndicate feature**

**Figure 18: Usefulness of nIntel charting feature**

**Figure 19: Usefulness of WordLearn HQ educational portal**

**Figure 20: Usefulness of nTranslate feature**

**Figure 21: Most useful feature(s) of nTrader**

# BIBLIOGRAPHY

Amazon. (2012). Amazon EC2 pricing. Retrieved January 10, 2013, from http://aws.amazon.com/ec2/pricing/

Federal Research Division, Library of Congress. (2011, December 30). Financial literacy among retail investors in the United States. Retrieved January 8, 2013, from http://www.sec.gov/news/studies/2012/917-financial-literacy-study-part2.pdf

Foxx, A. (n.d.). Infographic: Social games on Facebook and Google+ what makes them so popular? Retrieved 01 08, 2013, from http://www.prismcasino.com/entertainment/what-makes-social-games-on-facebook-and-google-so-popular-2779/

Moneysense. (2005, July). National financial literacy survey 2005 findings and recommendation. Retrieved January 9, 2013, from http://www.mas.gov.sg/~/media/resource/news\_room/press\_releases/2005/Summary%20Report.ashx

Property Guru. (2012, July 25). Singapore is 2nd most financially literate country in Asia Pacific. Retrieved January 8, 2013, from http://sg.news.yahoo.com/singapore-2nd-most-financial-literate-country-asia-pacific-055931832--sector.html

SiliconIndia. (2012, April 26). 5 mistakes young investors make. Retrieved January 9, 2013, from http://www.siliconindia.com/finance/news/5-Mistakes-Young-Investors-Make-nid-113656.html

TechTarget. (2012, April 24). Top 10 cloud computing providers of 2012. Retrieved January 08, 2013, from http://searchcloudcomputing.techtarget.com/photostory/2240149049/Top-10-cloud-providers-of-2012/11/1-Amazon-Web-Services#contentCompress

The Echo Boom. (2010, September 29). The echo boom statistics. Retrieved January 9, 2013, from http://theechoboom.com/2010/09/the-echo-boom-statistics/

The Wall Street Journal. (2012, February 5). Gen Y: The kids are all right. Retrieved January 9, 2012, from http://online.wsj.com/article/SB10001424052970204652904577192761903962328.html