FINANCIAL VISUALIZATION APPLICATION ADOPTING BIMODAL VISUALIZATION

by

Noor Magdelina binti Mohamad Dahalan

Dissertation submitted in partial fulfillment of
the requirement for the
Bachelor of Technology (Hons)
(Business Information Systems)

JANUARY 2008

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Financial Visualization Application adopting Bimodal Visualization

by

Noor Magdelina bt Mohamad Dahalan

A project dissertation submitted to the

Business Information Systems Programme

Universiti Teknologi PETRONAS

in partial fulfilment of the requirement for the

BACHELOR OF TECHNOLOGY (Hons)

(BUSINESS INFORMATION SYSTEMS)

Approved by,

_____

(Dr. Mohamed Nordin Zakaria)

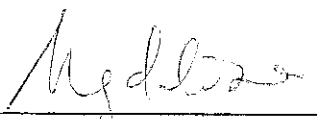
UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

January 2008

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own concept as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.



(NOOR MAGDELINABINTI MOHAMAD DAHALAN)

ACKNOWLEDGEMENT

First of all thanks to the Merciful full God of Allah to allow me to finish this project, without His bless and barakah this project should unable to reach the current progress and accomplish. All the devotion and effort were dedicating to my family especially my parents, Mohamad Dahalan and Zaleha Sobeng for their support to encourage me in many ways to stay focus and reach goal of the project and other study as complete the degree title.

Thank you to cooperative lecturers of Universiti Teknologi PETRONAS for their support and part of main contributor in this project. To name it few here, started with Dr Mohamed Nordin Zakaria as the most supportive and optimistic supervisor and also Ms. Emy Elyanee, Dr. P.D Dominic, Ms Shakirah, Mrs. Nazleeni, Mr Yew Khang Hooi, Mrs. Rohiza Ahmad, and Dr Wan Fatimah who also supports and gave resources and critics on the project.

Thank you also to classmates and friends who please to lend their ear to listen on draft idea of the project and spare their time for critics and supports during project research and implementation. Lastly, thanks again to all the people who is involved in this project directly or indirectly and thank you for all the efforts which contributed that are really appreciated.

ABSTRACT

The visualization system considers financial people's needs and approaches alternative way to understand financial data. This is one of the processes of analyzing and converting data from two different modalities into graphics. This allows financial decision makers and financial analyst to gain insight into the data, draw conclusions and directly interact with the data. The purpose of this project is to develop visualization for financial instruments – equity and to research the effectiveness of financial visualization for financial trader and investors. To make sure financial visualization for equities will work, this project will focus first in other element from finance which is share price. The reason to choose share price is because share price is one of important financial elements in financial market. This visualization system deals with two modalities of information; numerical and textual.

Financial trader or investor make his/her decision based on the behaviour of equities or share price over a certain period of time and consults other sources of information directly or indirectly with the instruments. These sources include internal factor of certain industry or economies and about the sector and wider issues that may affect on the instrument which equities.

ABBREVIATION (S)

GIDA	Generic Information – based Decision Assistant
FTSE	Financial Trade Stock Exchange
SATISFI	Sentiment and Time Series: Financial Analysis System
IDE	Integrated Development Environment
VB	Visual Basic
SDM	System Development Method
IT	Information Technology

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

This project will focus on developing an application that shows the correlation between time series of financial data and market sentiment using bimodal visualization and techniques. This visualization system deals with two modalities of information such as numerical and textual.

The available information in financial markets mainly comprises market instruments' indices, market reports, market news and announcements. Two main modalities are used in communication related to the financial markets. There are sequences of numerical data indexes by time, including number of shares traded, their values, and the values of the aggregates. The second modality is written language such as financial reports including market sentiment, other stories affected by and affecting the financial markets. All this data is not as discretely time indexed.

Data visualization methods are used quite extensively by the vendors of financial news. Different techniques were also developed for the same purpose like breakdown visualization, which is a visual spreadsheet format for comparison of adjacent visualizations.

Visualization techniques depend on only one modality of information, which is numeric. A problem arises when a different modality like textual information is introduction. Text usually displayed as stream of characters running under the time series display for much of the financial news is about change the value of instruments.

1.2 PROBLEM STATEMENT

Currently, visualization for financial data using only one modality of information which is "numeric" has been developed. A new problem arises when different modality like "textual" information been introduced. The financial trader receives market information in at least two forms: numbers and text. While various types of chart can be used to display numerical data in a form that is easy to digest and analyse, the financial trader is expected to read literally hundreds of stories. Even if this only means reading the headlines or skim-reading, this is still a monumental task. Surveys of financial traders show that rather than using informative, complex techniques, simple visualisation techniques were always preferred. Clearly, a system that can simultaneously deal with numeric and textual data would be of benefit.

1.3 OBJECTIVE

- To visualize potential change in the mood of market
- To visualize the data and information contents in two different modalities
- To develop an application that can manage a collection of news stories and historical time-series
- To analyze the news stories for market sentiment
- To correlate financial time-series with a time-series of keywords describing the positive and negative mood of the market

1.4 SCOPE

The application will focus on developing a prototype that correlates financial time-series together with keywords that describing up and down of current market. Despite the fact that textual information is not as discretely time indexed, this prototype demonstrates how news reports can be organised and analysed in order to assess and quantify the general mood of the trading market. The assessment is about the sentiment, which is conveyed in an individual news item: does the news item express a positive sentiment or a negative sentiment? Such expression is deeply embedded in language. Nevertheless, one can argue that the occurrence of words related to progress, profits, rising share prices may boost traders' confidence and the opposite, traumatic accidents, losses, falling share prices, will dent the traders' confidence.

The application is focus on help financial trader making decision on interpreting buy/sell signal from the visualization.

CHAPTER 2

LITERATURE REVIEW

2.1 BIMODAL VISUALIZATION – A FINANCIAL TRADING CASE STUDY

- Current application – prediction in financial trading
- Visualization for Financial Market
  - Introduce "Charting" technique to generate decision-making
- Generate buy/sell signals for financial instruments
- Visualizing Market Sentiment
  - indicate how traders feel about market – use a few term

2.1.1 System SATISFI

- A visualization system for viewing a text corpus, queried according to the various attributes of its constituent text has been developing in Java as a key component.
- Work with Reuter's country and industry sector categories together with Reuters supplied keywords.

2.1.2 Correlation Visualization adopting Bimodal Visualization – Current Project

- Can view financial news in Malaysia and about particular sector
- The visualization system work with The Edge Daily in Malaysia's market analysis.
- Correlation Visualization was developed to show correlation between time series and market sentiment.

2.2 NetBeans

NetBeans refers to both a platform for the development of Java desktop applications, and an integrated development environment (IDE) developed using the NetBeans Platform. The NetBeans Platform allows applications to be developed from a set of modular software components called *modules*. A module is a Java archive file that contains Java classes written to interact with the NetBeans Open APIs and a manifest file that identifies it as a module. Applications built on modules can be extended by adding new modules. Since modules can be developed independently, applications based on the NetBeans platform can be easily and powerfully extended by third party developers.

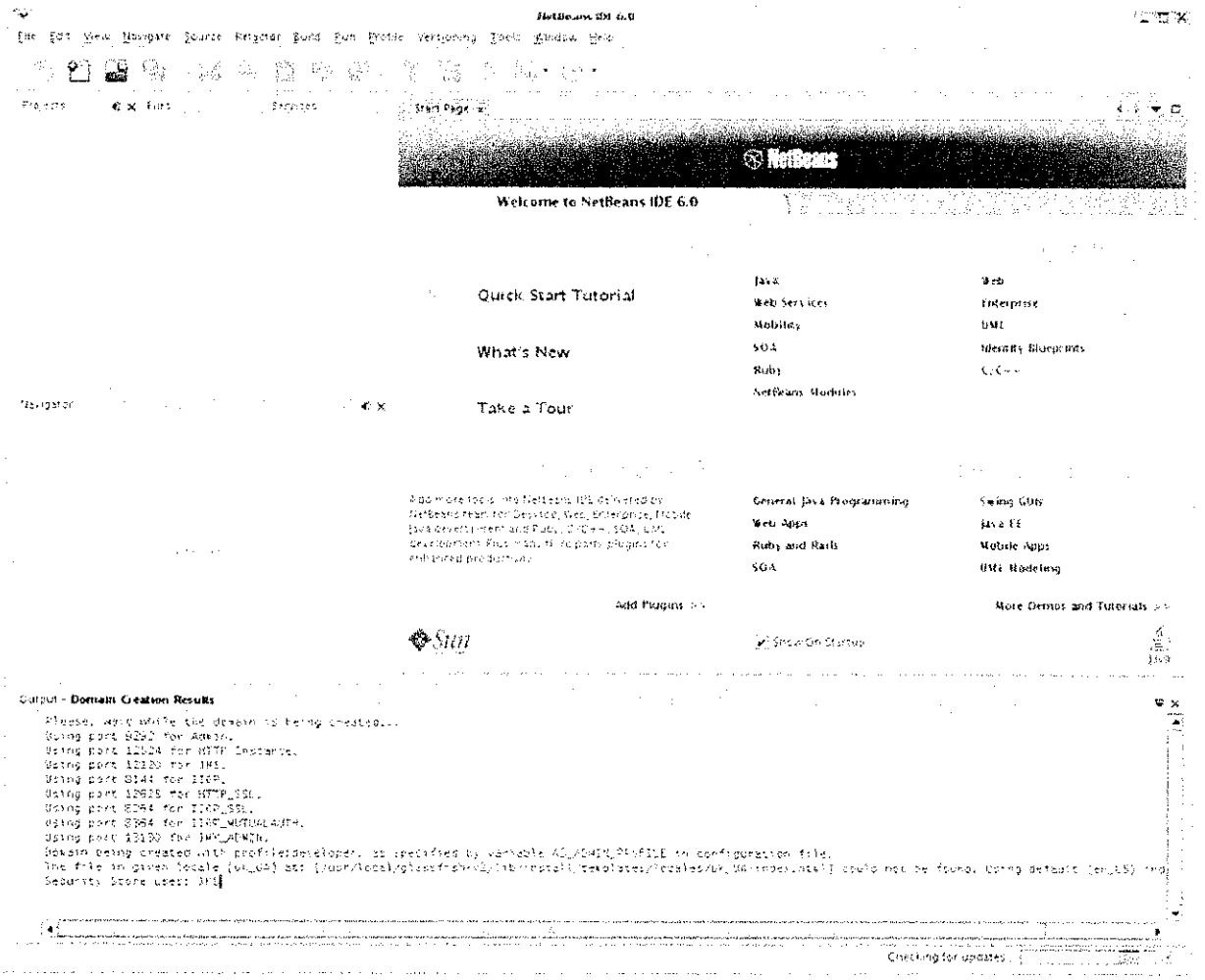


Figure 1

2.2.1 NetBeans Platform

The NetBeans Platform is a reusable framework for simplifying the development of other desktop applications. When an application based on the NetBeans Platform is run, the platform's main class is executed. Available modules are located, placed in an in-memory registry, and the modules' start-up tasks are executed. Generally, a module's code is loaded into memory only as it is needed.

Applications can install modules dynamically. Any application can include the Update Centre module to allow users of the application to download digitally-signed upgrades and new features directly into the running application. Reinstalling an upgrade or a new release does not force users to download the entire application again.

The platform offers services common to desktop applications, allowing developers to focus on the logic specific to their application. Among the features of the platform are:

- User interface management (e.g. menus and toolbars)
- User settings management
- Storage management (saving and loading any kind of data)
- Window management
- Wizard framework (supports step-by-step dialogs)

2.2.2 NetBeans IDE

The NetBeans IDE is an open-source integrated development environment written entirely in Java using the NetBeans Platform. NetBeans IDE supports development of all Java application types (J2SE, web, EJB and mobile applications) out of the box. Among other features are an Ant-based project system, version control and refactoring.

The current version is NetBeans IDE 6.0, which was released in December 2007. NetBeans IDE 6.0 extends the existing Java EE features (including Java Persistence support, EJB 3 and JAX-WS). Additionally, the NetBeans Enterprise Pack supports development of

Java EE 5 enterprise applications, including SOA visual design tools, XML schema tools, web services orchestration (for BPEL), and UML modelling. The NetBeans C/C++ Pack supports C/C++ projects. NetBeans 5.5.1 builds on the functionality of NetBeans 5.5 and also provides several bug fixes.

NetBeans IDE 6.0 builds upon the previous version 5.5.1, which introduced comprehensive support for developing IDE modules and rich client applications based on the NetBeans platform, a new GUI builder (formerly known as "Project Matisse"), new and redesigned CVS support, Weblogic 9 and JBoss 4 support, and many editor enhancements.

2.2.3 Modularity

All the functions of the IDE are provided by modules. Each module provides a well defined function, such as support for the Java language, editing, or support for the CVS versioning system. NetBeans contains all the modules needed for Java development in a single download, allowing the user to start working immediately. Modules also allow NetBeans to be extended. New features, such as support for other programming languages, can be added by installing additional modules. For instance, Sun Studio, Sun Java Studio Enterprise, and Sun Java Studio Creator from Sun Microsystems are all based on the NetBeans IDE.

2.3 Visual Basic (Microsoft Visual Studio 2005)

At early phase of this project, I decided to use Java but in the end I found Java is more challenging and difficult for me to developing this application. After make a few research through out the web and books, I decide to develop this application using Visual Basic.

2.4 The Edge Daily Malaysia

TheEdgeDaily is the financial and investment daily news website of The Edge weekly. It brings unprecedented advantage for people in search of business and investment news on Malaysia and the region. Our readers will have access to breaking news throughout the day, every trading day – an edge they need to stay ahead of today's fast paced market moves.

TheEdgeDaily is brought to you by The Edge Communications Sdn Bhd, which publishes The Edge, Malaysia's best selling and highly regarded publication on business and investment. A complete, separate dedicated editorial team of writers and reporters work extensively to provide the latest on business and investment news to help our readers make informed investment decisions.

2.5 Microsoft Access and Microsoft Excel

In developing this project, Microsoft Access and Excel is important and needed software. The system using Microsoft Access to store and retrieve data that later subsequently will generated correlation visualization automatically. Microsoft Excel is using to display the generated correlation visualization or graph.

CHAPTER 3

METHODOLOGY

3.1 SPIRAL MODEL

The spiral model is a software development process combining elements of both design and prototyping-in-stages, in an effort to combine advantages of top-down and bottom-up concepts. Also known as the spiral lifecycle model, it is a systems development method (SDM) used in information technology (IT). This model of development combines the features of the prototyping model and the waterfall model. The spiral model is intended for large, expensive and complicated projects.

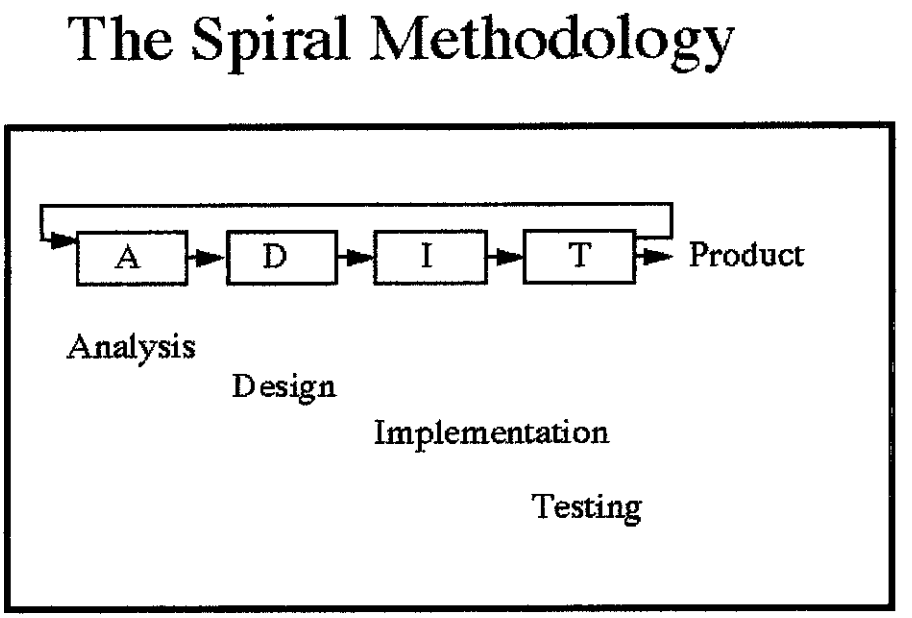


Figure 2

3.2 PROTOTYPING METHOD

Throughout January until March 2008 I have try to used spiral as my method for this project, I have encountered many problems in developing the prototype. As for conclusion I decided to change my method from spiral to prototype method. As far as I'm concern, this project still in design phase. I have problems in writing code to create graph for my application.

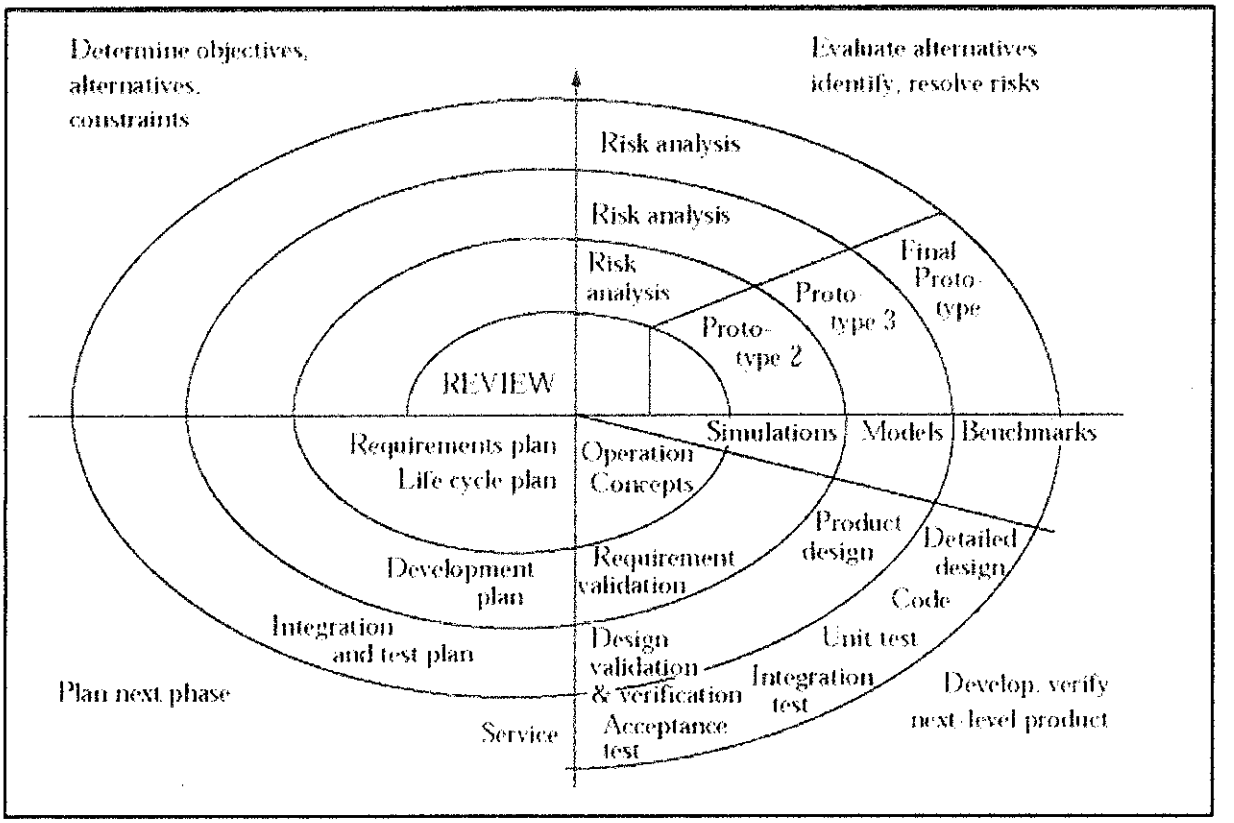


Figure 3

3.2.1 Current Phase

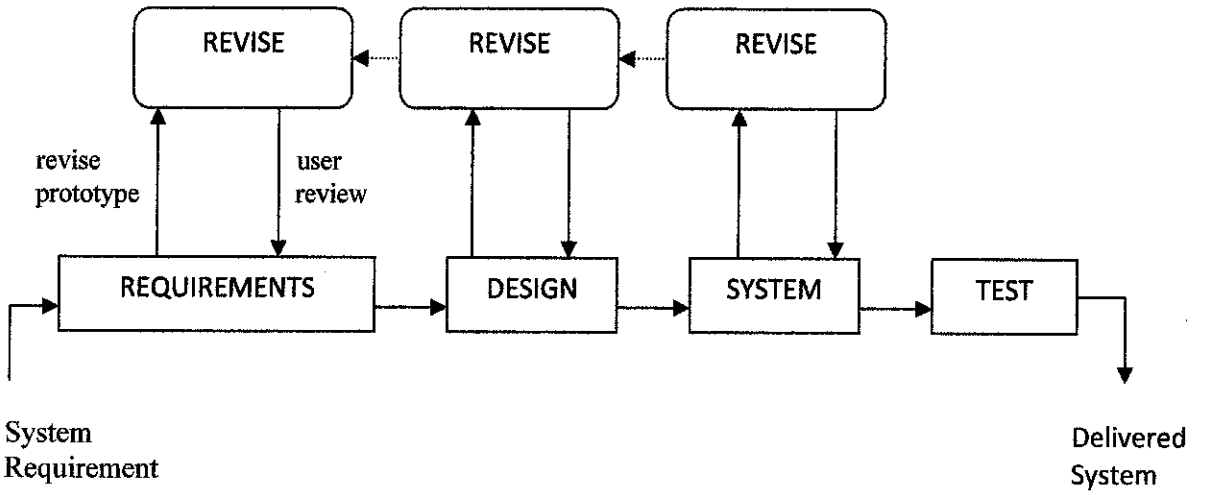


Figure 4

Prototyping allows all or part of a system to be constructed quickly to understand and clarify issues. The system requirement or design requires repeated investigation to ensure user have common understanding what is needed and what is proposed. Overall is, this method focus on reducing risk and uncertainty in system development.

3.3 WORK FLOW

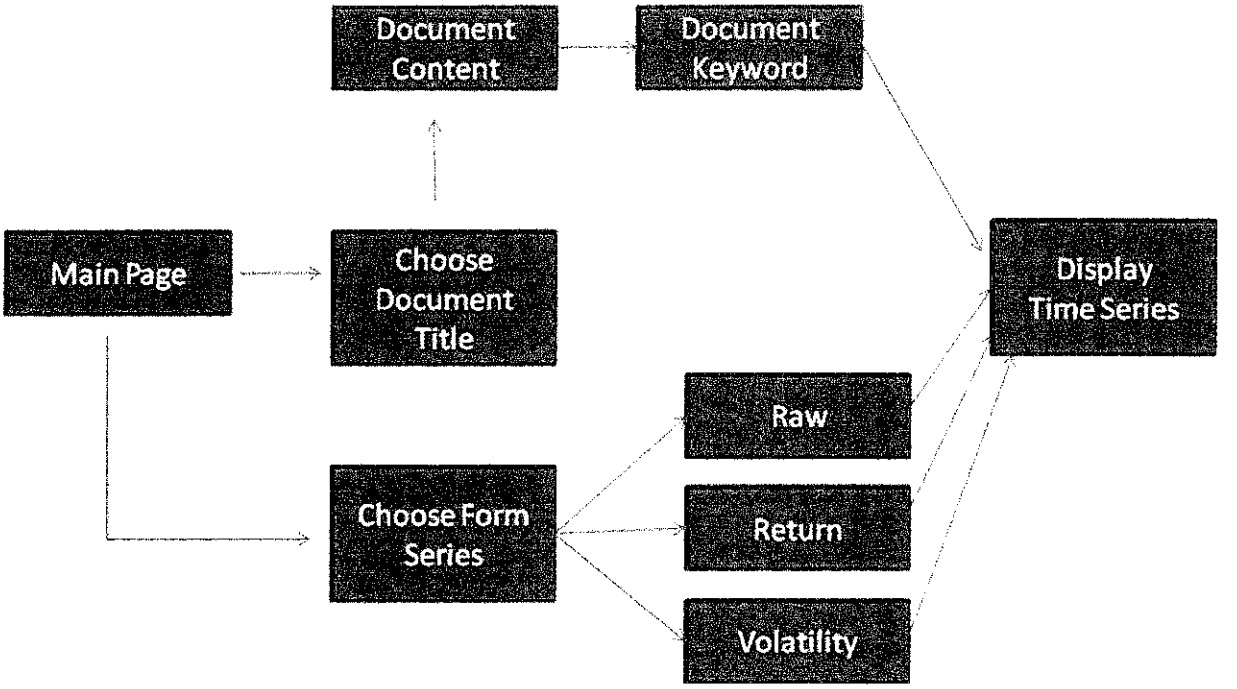


Figure 5

3.4 ACTIVITY DIAGRAM/SYSTEM PROCESS

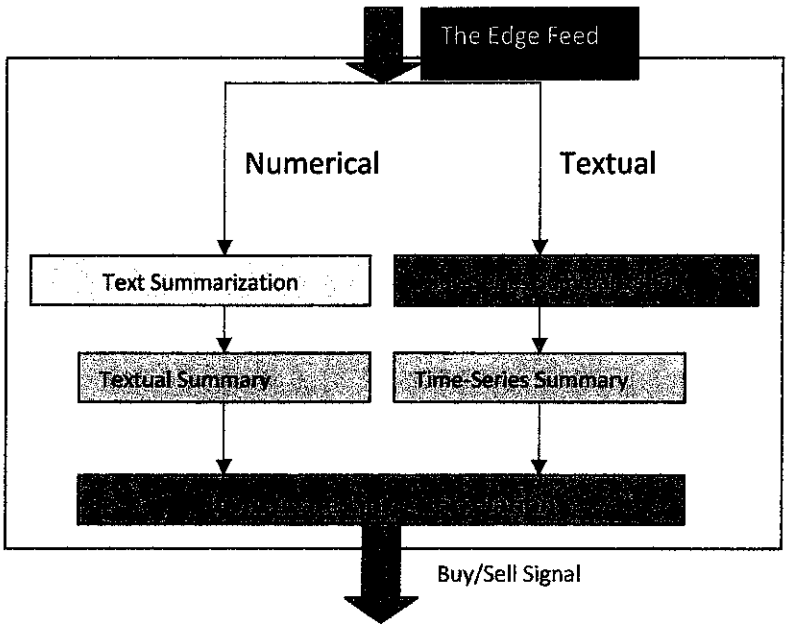


Figure 6

This figure outlines basic approach that is adopted to process of bimodal information. The source of data is from The Edge Daily Malaysia and has been analyze to generate unique keywords. This framework:

a) Summarizes financial news articles using linguistic resources

b) Summarizes locally volatile financial time series in terms of the turning points

c) Facilitates the visualization of the buy/sell signal. This will define text summarization and time series summarization modules in greater details.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 FIRST PHASE

4.1.1 Draw X-Line Axis and Y-Line Axis

A visualization system for viewing a text corpus, queried according to the various attributes of its constituent texts can be developing using Java and to be more specific is using JCreator and NetBeans software. The system is plan to be writing in Java and can be accessible on the world-wide web. The visualization system can work with financial news and industry sector with supplied keywords. The user can view news about a particular country of sector.

The visualization system on the web shows weekly news and display content of an individual news item together with the keywords associated with the content. The visualization system can work with financial news categorisation. This system was developed to show the correlation between a time-series of an instrument is seldom used in its raw form rather derived forms are used: *return* and *volatility*. The index *return* is the logarithmic difference between two consecutive values and *volatility* is a measure of how fast or slowly the instrument moves *up* or *down* in terms of its value.

4.2 SECOND PHASE

The system has four major components that should have in the system:

4.2.1 Time Series Display:

The system can display three time series at a time. These time series comprise index values, upward movement indicators and downward indicators. Upward and downward movement indicators are the quantification of the market sentiment expresses in financial news. Over 70 terms each have been identified for conveying 'good' and 'bad' news. The movement indicator time series are synthesized by counting these movement indicator terms within the financial news published for a particular day. Each time series is normalised for proper display purposes.

4.2.2 Time Series Correlation:

Correlation is a measure of the degree of linear relationship between two time series. The system provides the user the facility of cross correlating two series in any form. Any series can be shifted forward or backward and cross correlation recalculated to determine whether the market is followed by the news or vice versa.

4.2.3 Document Display

This comprises two parts:

a. **Document Titles**: Clicking on any of the time series, displays the corresponding date's news titles.
b. **Document Content**: The content of any document title can be viewed by clicking that news title.

4.2.4 Document Analysis:

Whenever a document title is selected from the news list, the extracted sentiment keywords along with the frequencies are displayed in "Document Keyword" area. Positive sentiment keyword analysis details appear under the title of "Upward

Movement Indicators" and negative sentiment keyword analysis details appear under the title of "Downward Movement Indicators".

4.3 THIRD PHASE

After change from Java language to Visual Basic, I already design the interface of the application/system. I have already design three (3) out of four (4) major parts that should have in this system.

4.3.1 Interface Screen Shot (First Phase)

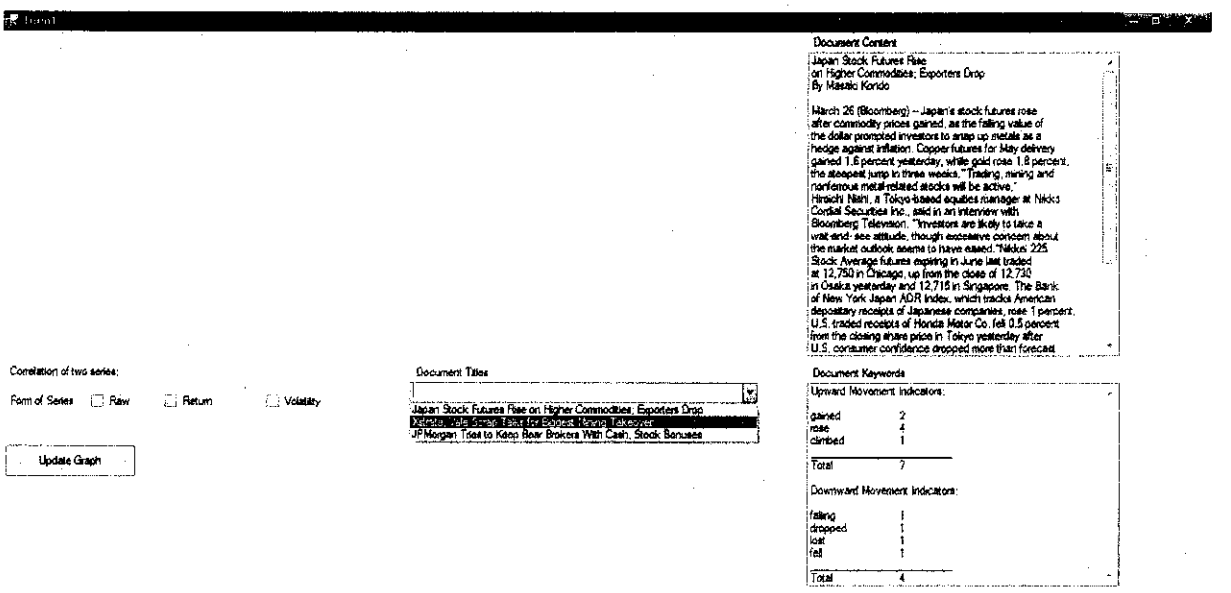


Figure 7

In Figure 7, we can see that three major parts of the system have been design and not fully integrated with coding. End user can select any financial news from document title to be read. The selected document will be display inside document content and will be automatically retrieve reserve word like gain, fall, rose and etc. which is a unique keywords indicator. These keywords later will be display inside document keyword part and will sum up the total number of upward and downward movement indicator. These will determine the

mood of market on that day. Later, it will correlate financial time-series with time-series of unique keyword to be display in chart.

4.3.2 Interface Screen Shot (Second Phase)

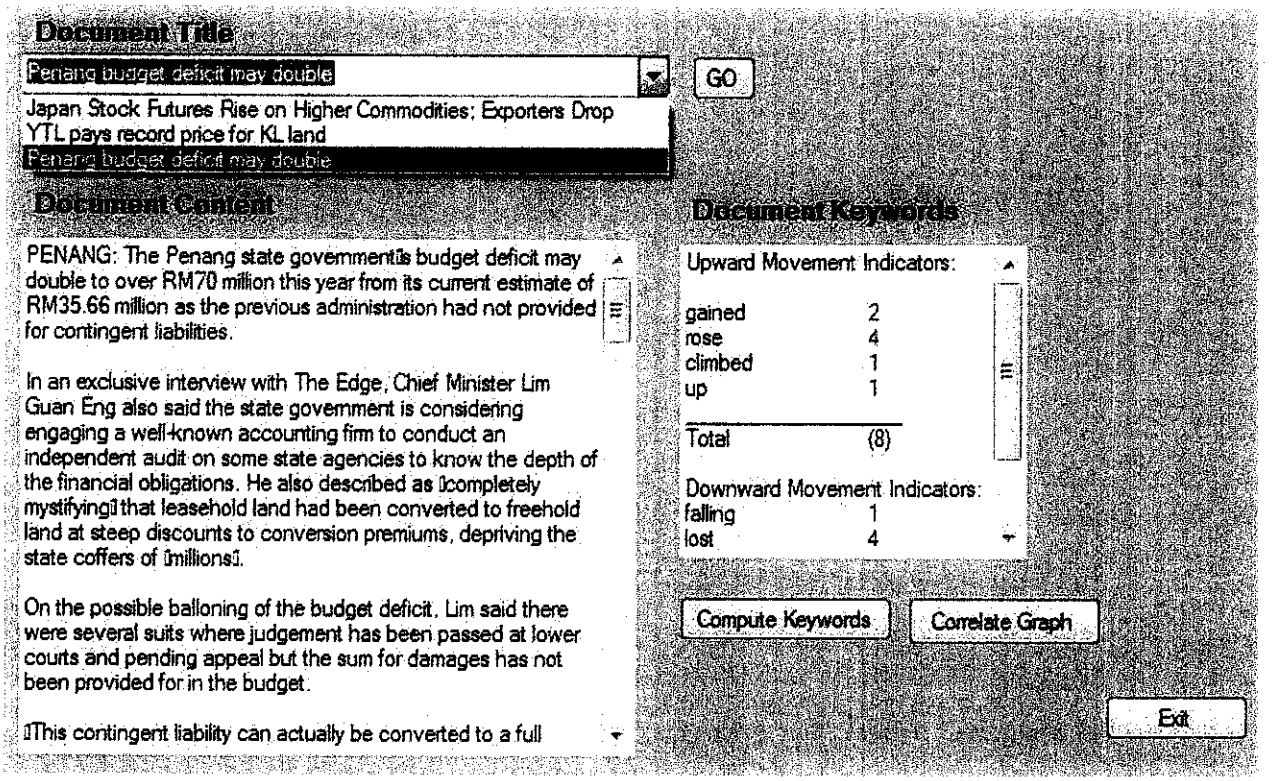


Figure 8

Figure 8 shows improvement of the system in function coordination and colour. The system also retrieves a few news stories from The Edge Daily Malaysia.

4.3.3 Interface Screen Shot (Third Phase)

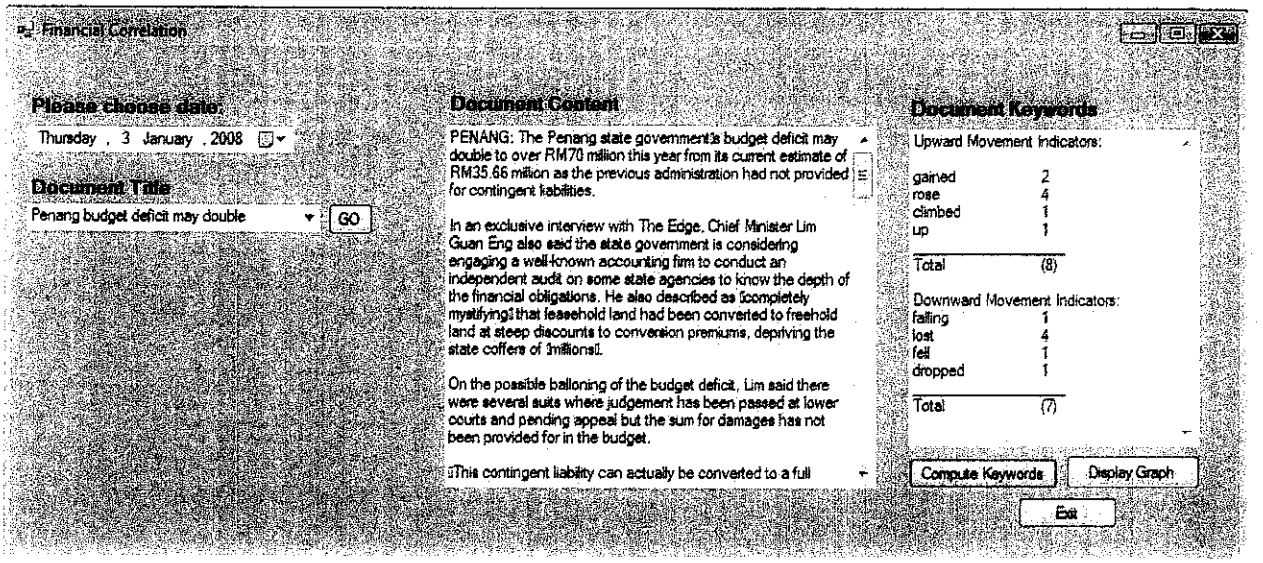


Figure 9

In third phase, a few functions have been added such as date. This is to ease user in choosing which date of news stories they want to read.

4.3.4 Numerical Visualization – Possible Result

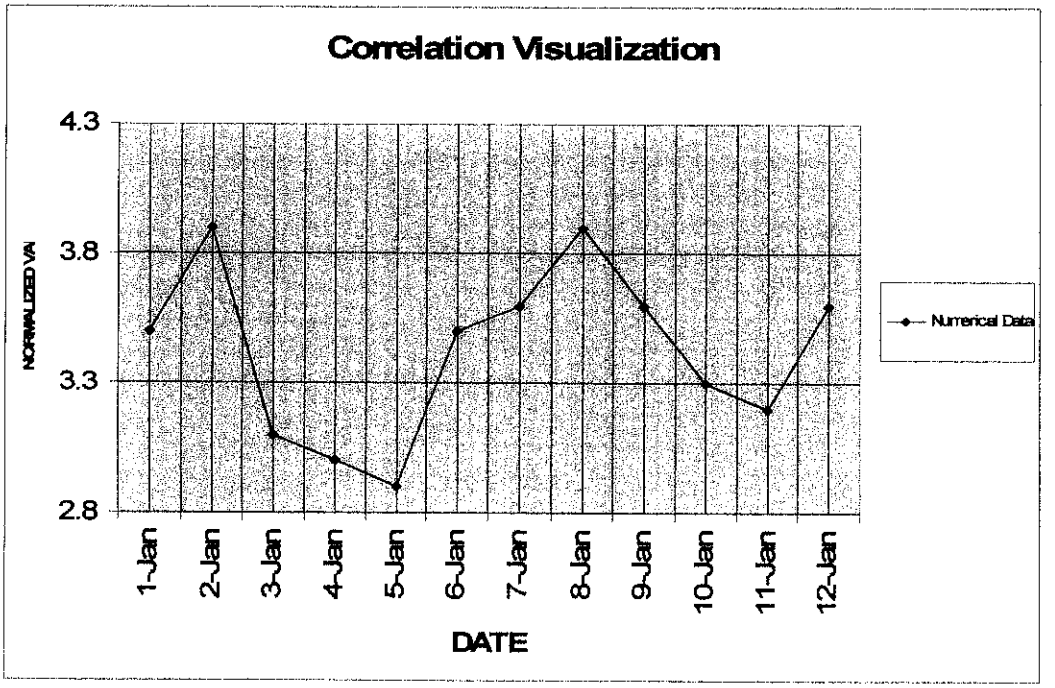


Figure 10

4.3.5 Textual Visualization

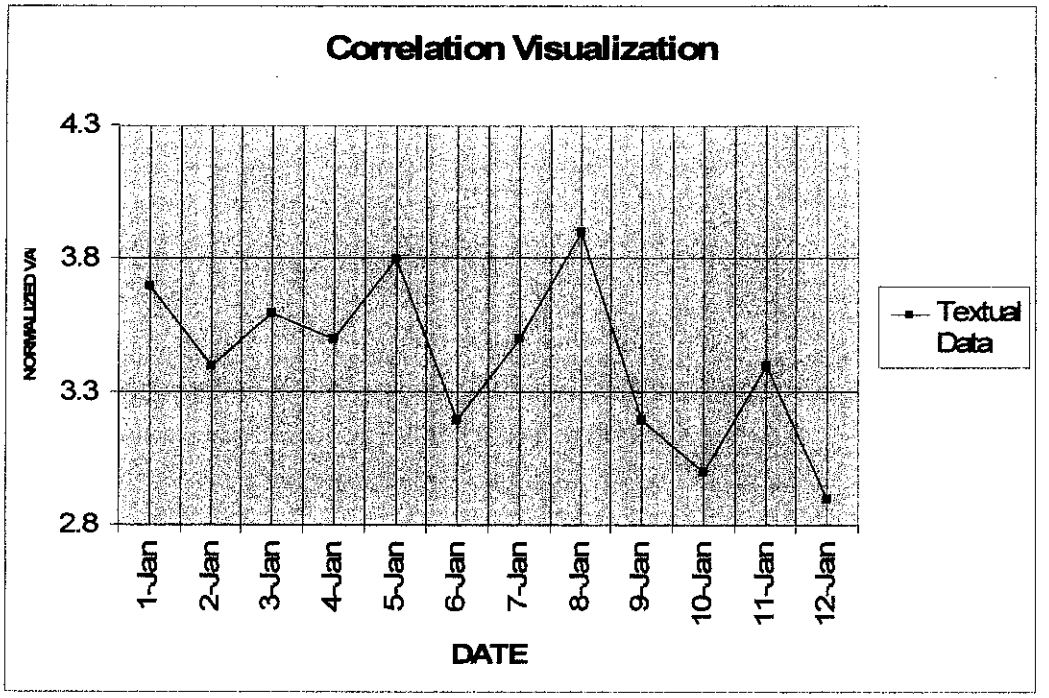


Figure 11

4.3.6 Correlation Visualization – Possible Result

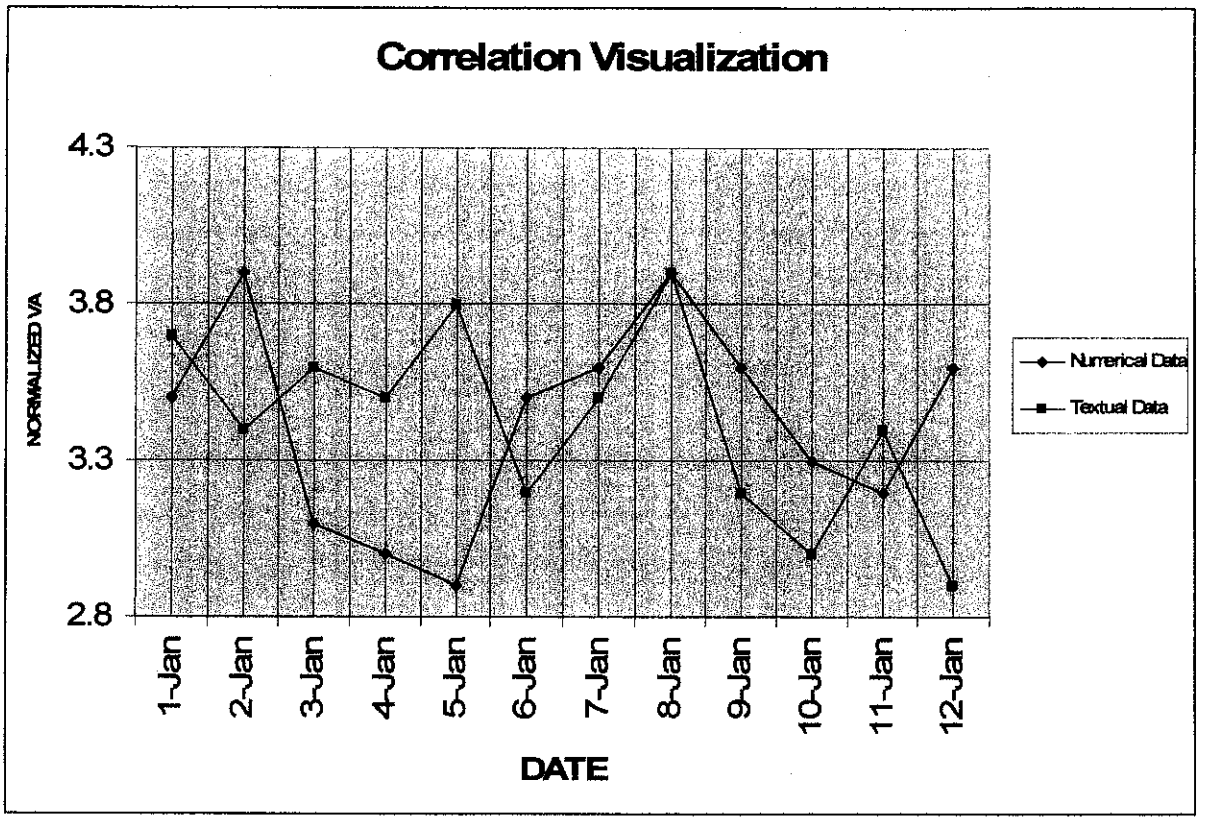


Figure 12

All the possible result above is generated using Microsoft Excel. The data available should be retrieving automatically but because the system is not currently running and online yet, all these data was hard-coded. In present, the system still not is able to automatically correlate between financial time-series with time-series of keywords.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

Visualization techniques depend on only one modality of information which is numeric. A new problem arises when a different modality like textual information is introduced. Text usually displayed as stream of characters running under the time series display, for much of the financial news is about change the value of instruments.

It shows how it is possible to visualize the data and information content in two different modalities. This visualization introduce different technique that able to receive and display a numerical time series of financial data, and its ability to correlate the series with an index of market sentiment should be its notable feature.

5.2 RECOMMENDATION

- In future, the system or application will be able to show how it is possible to visualize and correlate data and information contents in two different modalities.
- The system should be able to describe mood of the market only based on this correlation visualization.
- Hence, the system can be able to find out which of the turning points in the value of an instruments (from down to up or up to down) were caused by a specific news item using the visualization
- The system also possible to prove it as a prediction system, that it is should be used in investigating past behaviour of the market.

REFERENCES

- http://en.wikipedia.org/wiki/NetBeans
- The Edge Daily Malaysia : www.theedgedaily.com
- http://en.wikipedia.org/wiki/Spiral_model
- Bimodal Visualization: A Financial Trading Case Study, *Tugba Taskaya* and *Khurshid Ahmad*, 2003
- J.Saltz, J.Steinbach. "Wayfinding in Layered Information Worlds", CODATA Euro-American Workshop Visualization of Information and Data: Where We Are and Where Do We Go From Here?, 24-25 June 1997, pp. 6-7.
- K. Ahmad, D. Cheng, T. Taskaya, S. Ahmad, L. Gillam, P. Manomaisupat, H. Traboulsi, A. Hippisley, "The mood of the (financial) markets: In a corpus of words and of pictures", Proceedings of Corpus Linguistic 2003. Lancaster University, UK. http://www.computing.surrey.ac.uk/iemmg/publications.html
- Saif Ahmad, Paulo C F de Oliveira and Khurshid Ahmad, "Summarization of Multimodal Information", Department of Computing, University of Surrey Guildford, Surrey, GU2 7XH, UK

APPENDICES

Bimodal Visualisation: A Financial Trading Case Study

Tugba Taskaya and Khurshid Ahmad
Department of Computing
University of Surrey, Guildford, Surrey. GU2 7XH. UK
{ t.taskaya, k.ahmad }@surrey.ac.uk

Abstract

A visualisation system that deals with two modalities of information – numerical and textual – is presented. The current application domain is that of prediction in financial trading. The system synthesises news stories and time series data to generate buy/sell signals for financial instruments.

1. Introduction

Data and information visualisation systems are used in noisy environments that typically involve data streams encoded in a single modality of communication. Numerical data, organised as time-serial data, is used to study the behaviour of mobile phone owners or that of purchasing decisions of consumers; textual data is used to create the so-called knowledge maps, thematic maps and semantic maps showing preferential citation patterns and/or patterns of (key) word usages amongst scientists and technologists. Image data streams, whether still or video images, are analysed purely on the visual content of the images, to categorise and identify events or objects.

Usually, the analysis of the unimodal data helps in creating categories of people using mobile phones, buying one soap powder brand in preference to another, or categories of scientists focused on one scientific endeavour rather than another [1].

The data and information visualisation systems, working typically on one mode of communication modality, facilitate decision making for the human users of such systems. The human users then use other data sources, perhaps encoded in modalities of communication other than those used by the systems, to actually make their final decision. A good example is that of a financial trader dealing in financial instruments. The instruments includes shares, currencies, derivatives; the aggregate values of share within a national stock exchange or within an industrial/service/commercial sector is frequently used as a measure of the financial health and economic robustness of a national economy or that of the sector. The financial trader usually has access to a high-volume data stream that tells the trader how many shares were traded at a certain value, for example in intervals as small as a few seconds. The aggregates, like the Financial Times-Stock Exchange Indices, DowJones Industrial Averages, are displayed simultaneously. Essentially, the trader has access to a time series of numbers, which is analysed using well-established statistical techniques that project the quantity and value of the instrument at some future time.

Trend analysis is a critical task in financial-trading. A mixture of empirical and theory-based techniques has been developed to display the behaviour of financial (instrument) time series. Empirically, the behaviour of the time series is studied over a period of time (few hours, days or months), some of these patterns have a characteristic shape; a shape associated with a predictable behaviour at the end of the window. Patterns are called *head and shoulders* (two minima comprising a peak) or *harami* patterns [3]. Theoretical techniques are based on statistical properties of time series: autoregression, Fourier decomposition, underlying randomness, modelled by established statistical distributions. Many of these techniques originate in digital signal processing.

The trader makes his or her buy/sell decision based on the behaviour of the instrument over a certain period of time and consults other sources of information directly or indirectly connected with the instrument: these sources include news about the national economy or economies, about the sector and wider issues that may have catastrophic or beneficial effect on the on the instrument.

A range of data and information visualisation techniques have been developed to display the high volume series: data visualisation includes the so-called Japanese candle stick patterns where for each point in time the *candle stick* shows high/low and current value

dings of the Seventh International Conference on Information Visualization (IV'03)
547/03 $17.00 © 2003 IEEE

of the instrument together with information as to whether the instrument is falling or rising in value.

The display of news streams has been made very user friendly and the emergence of an XML-based standard, called newsML, means that each news report can be marked up almost at the content level [4]. News streams can be according to the user profiles. This makes it easier filtered in or out for the end user to selectively receive what he or she wants from 2000 or so daily news stories typically sent by a single news agency like Reuters. However, the financial trader is expected to read literally hundreds of stories, either by just reading the headline, or by skimming or reading in detail.

Surveys of financial traders showed that rather than using informative, complex techniques, simple visualisation techniques were always preferred [2].

News agency also facilitate search for past news reports literally over decades. The indexing of news items is generally good but it is largely manual indexing by keywords and categories available to the human sub-editor who in effect processes the news for it to go on the news wire. The marked-up news, for keywords and categories (relating to different instrument and market sector) is in principle, an attempt to visualise the content of the individual news reports.

To summarise then, the financial trader receives market information in at least two modalities: numerical, as in time serial data related to one or more instruments, and textual, as in news wires about the instrument directly or indirectly. The trader then synthesises the information in the time series and the news stream before making the buy and sell decision. The above scenario, in which the financial traders find themselves in on a daily basis, can benefit from a system that can simultaneously deal with the two principal modalities, numeric and textual. Each of the modalities can be processed separately, and the results of the analysis then be synthesised to generate a buy and sell decision. It is the prototype of such a system that we describe briefly. This prototype demonstrates how news reports can be organised and analysed in order to assess and quantify the general mood of the trading market. The assessment is about the sentiment, which is conveyed in individual news item: does the news item express a positive sentiment or a negative sentiment? Such expression is deeply embedded in language. Nevertheless, one can argue that the preponderance of words related to progress, profits, rising share prices may boost traders' confidence and the opposite, traumatic accidents, losses, falling share prices, will dent the traders' confidence.

Section 2 is analysing visualisation methods used in financial information.

Section 3 comprises an outline of a prototype system being developed for an EU 5th Framework project, GIDA. The prototype can manage a collection of news stories – called a corpus of texts – and historical time series; can analyse the news stories for the market sentiment (boost/depression); can correlate financial time series with a time series of keywords describing the positive mood of the market and other time series describing the negative mood.

Section 4 describes the system in operation where it correlates a time series of an instrument (FTSE – 100) with mainly UK-related news to generate a buy and sell signal. Section 5 concludes the paper.

2. Visualisation for Financial Markets

The available information in financial markets mainly comprises market instruments' indices, market reports, market news and announcements. Two main modalities are used in communication related to the financial markets. First, there are sequences of numerical data indexed by time, including volumes of shares traded, their values, and the values of the aggregates, e.g. overall market aggregates (FTSE, Dow Jones), or differences in the aggregates (derivative). The second modality is that of (written) language: financial reports, including market sentiment, other stories affected by and affecting the financial markets. This data is not as discretely time indexed.

Charting is a standard technique used in financial decision-making. The set of techniques used in technical analysis helps to plot price movements, volume, settlement prices, open interest, and other indicators, in order to anticipate future price movements [5]. A financial time series is often visualised in three formats: Line Chart, Bar Chart, and Candle Chart [see Figure 1, Figure 2, Figure 3].

These charts are used by a number of financial web sites such as "Yahoo Finance"[6], "Prophet.Net"[7], "Financial Times" [8]. These websites also comprise:

(a) Financial news related to selected instrument;
(b) Fundamentals such as dividend, market capitalization;
(c) Statistical tools for analysis such as moving averages, bollinger bands, RSI; and,
(d) Options like time frames, historical prices.

The time intervals can also be changed by the end user. In addition to the charts used for displaying historic and current numerical values of an instrument, the financial news vendors provide access to the so-called fundamental financial details including dividends, market capitalisation and so on.

Data visualisation methods are used quite extensively by the vendors of financial news: Methods like Inselberg 's parallel coordinates [9], scatter plots are used to explore the relationships between prices for example sectors versus sector values such as net income, total assets. Different techniques were also

developed for the same purpose like breakdown visualisation, which is a visual spreadsheet format for comparison of adjacent visualizations [10].

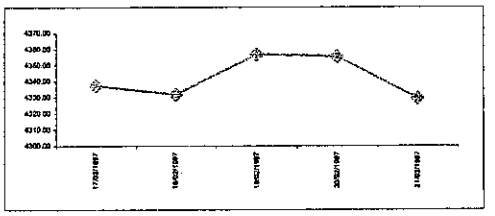


Figure 1. Line Chart uses dots versus time in which dots represents generally closing price.

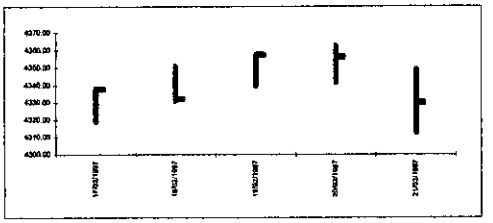


Figure 2. A three prolonged bar: right represents opening price; top the highest price and bottom is the lowest.

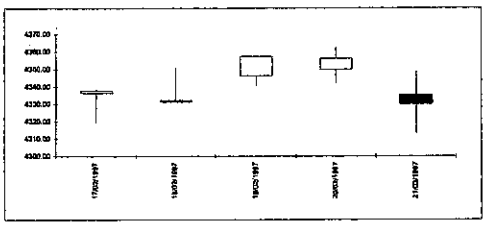


Figure 3. This chart can be used to display the high and the low value of the price along the vertical axis. The horizontal axis is used for time information and also to display whether or not the opening price at the beginning of the trading period was higher or lower than the closing price. A higher opening price and lower closing price is displayed by a solid rectangle and the opposite by a hollow rectangle.

Topographic financial maps generally consist of regions, which indicate sectors or funds. Companies that belong to either stock index or user's portfolio are browsed from the corresponding region. The area that the sector and company occupy is proportional to the sector price and the company price respectively [11], [12], [13]. User interaction with the map is supported generally in two ways. First a tool tip text with a summary of company information pops up whenever the mouse points to the area where that company is represented. And secondly, company information is browsed in detail whenever a mouse clicking action is performed.

Visualisation techniques discussed so far depend on only one modality of information, which is numeric. A new problem arises when a different modality like textual information is introduced.

Text usually displayed as a 'ticker tape' that is stream of characters running under the time series display, for much of the financial news is about change the value of instruments. Equally important is the 'indefinable' market sentiment: The potential change in the mood of the market –how is this to be visualised?

3. Visualising Market Sentiment

Reports about the financial markets sometimes use metaphorical terms to indicate how the traders feel about the market: terms like '*bear*' and '*bull*' markets are used to describe that traders are shy and reticent about buying stocks and *bull* suggest that their mood is to buy aggressively. There are a number of metaphors in addition to bear and bull. News report contains simpler expressions like:

FTSE set to fall
Royal and Sun hit by asbestos suit
Energy tax hurts business

Each of the underlined verbs *fall, hit, hurts* conveys a negative sentiment. On the other hand, verbs like *rise, jump,* and *climb* may convey a positive sentiment. It is true that a word in natural language may have a range of meanings: the token *rose* may mean the flower *rose*; it can be a name, as in Mr/Ms Rose in addition to meaning 'increase in the value of a share or currency'. However, in financial texts tokens like *rose, fall* have a restricted meaning [14]. This is particularly true when these words are used in the following contexts:

{Shares, currencies} *rose/fell* {by} X percent
{Shares, currency} *rose/fell* {to} X percent

One heuristic will be to use the frequency of these two verbs if and only if they occur in the context described above. The use of this heuristic allows us to quantify the market sentiment in as much as a computer system can ever evaluate any sentiment.

We have analysed a 2 million word corpus of text comprising financial news supplied by the Reuters News Agency for the calendar year 2000. These texts were analysed by our text analysis system, System Quirk (which performs text analysis for our prototype) and we extracted the frequency of the verbs *rose* and *fell*. These frequencies were then plotted together with the FTSE 100 index. Figure 1 shows how the sentiment time series correlate with FTSE.

We have performed the analysis over a larger corpus of Financial News by Reuters over 3-year period (2000-

2002), comprising over 10 million words, and found that the heuristics mentioned above to hold as well. In addition, many other sentiment verbs and prepositions (*up*, *down*, *adrift*) have been identified and used to compute the market sentiment. The extraction methodology is described in detail in [14].

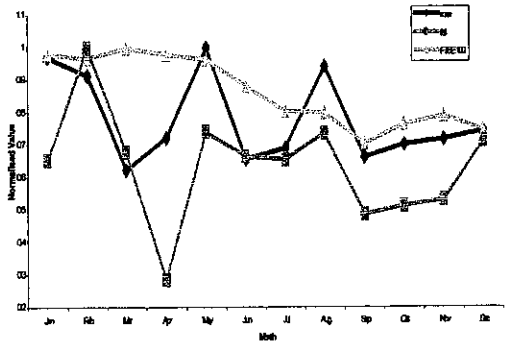


Figure 4. The sentiment time series correlate with the FTSE

4. A System for Visualising and Correlating Time Ordered Market Data

4.1. Visualising a News Corpus

Typically, a news agency like Reuters or Bloomberg will supply over 2000 stories per day comprising between 300-800 words. Reuters news stories are XML-formatted which give amongst other information, industry sector and country information.

Queries against a database of news stories can be handled provided the stories have been pre-indexed on to the date of publication, keywords characterising the content of the news, and, in our case, key verbs and prepositions characterising the sentiments, with positive and negative, expressed in and by each of news stories. Once a query has been successfully matched, it is important that the news story be displayed in full together with the frequency of the keyword, key verbs and prepositions.

A visualisation system for viewing a text corpus, queried according to the various attributes of its constituent texts, has been developed in Java as a key component to our financial trading prototype (SATISFI: "Sentiment and Time Series: Financial Analysis System"). The system is written in Java and is accessible on the world-wide web (see Figure 5).

Our visualisation system can work with Reuters country and industry sector categories together with Reuters supplied keywords. The user can view news about a particular country of sector, see the distribution of number of news stories over time soon (see Figure 6).

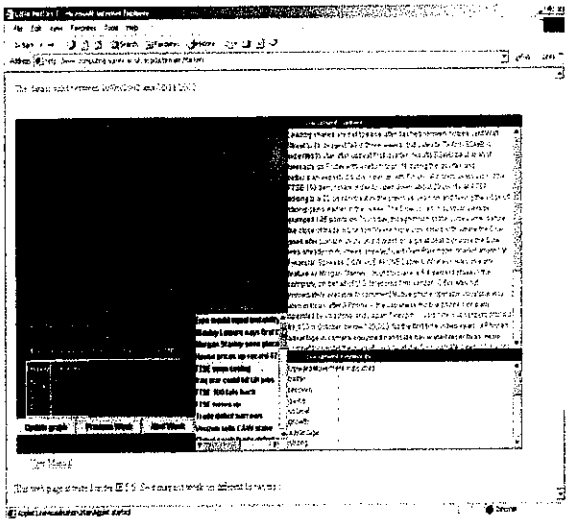


Figure 5. The visualisation system on the web shows weekly news. The system displays content of individual news items together with the keywords associated with the content.

4.2. System SATISFI-Visualising Correlation between Sentiment and Market Indices

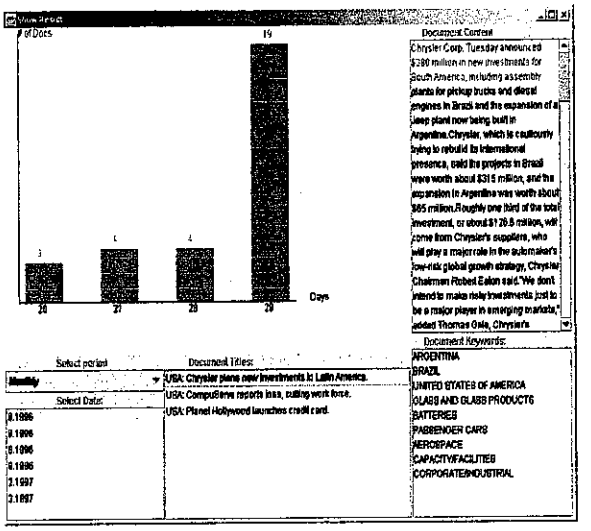


Figure 6. The visualisation system can work with Reuters news categorisation.

SATISFI was developed to show the correlation between a time series of an instrument and market sentiment. The time series of an instrument is seldom used in its raw form rather derived forms are used: *return* and *volatility*. The index *return* is the logarithmic difference between two consecutive values and

volatility is a measure of how fast or slowly the instrument moves *up* or *down* in terms of its value.

Figure 7 shows how the prototype system works in terms of its four components: each helping to visualise either the market sentiment or an instrument's time variation. The value of the correlation helps the user to make a buy/sell decision.

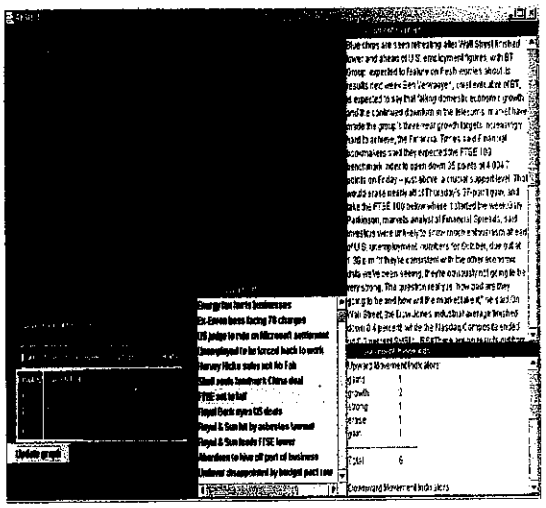


Figure 7. SATISFI prototype

SATISFI has four major components that have been fully integrated as shown in Figure 7.

I. **Time Series Display**: SATISFI can display three time series at a time. These time series comprise FTSE-100 close index values, *upward movement indicators* and *downward movement indicators*. As discussed above, upward and downward movement indicators are the quantification of the market sentiment expressed in financial news. Over 70 terms each have been identified for conveying 'good' and 'bad' news. For example upward movement indicators would contain terms like 'up, rise, growth' etc. while downward movement indicators would contain terms like 'down, fall' etc. The movement indicator time series are synthesized by counting these movement indicator terms within the financial news published for a particular day. Each time series is normalised for proper display purposes. SATISFI is capable of displaying the above time series in three forms:

   a. **Raw** form denotes the original time series.
   b. **Return** form refers to the logarithmic difference between two consecutive values.
   c. **Volatility** (historical volatility) is the relative rate at which the time series moves up or down.

II. **Time Series Correlation:** Correlation is a measure of the degree of linear relationship between two time series. SATISFI provides the user the facility of cross correlating two series in any form (raw, return, volatility). Any series can be shifted forward or backward and cross correlation recalculated to determine whether the market is followed by the news or vice versa.

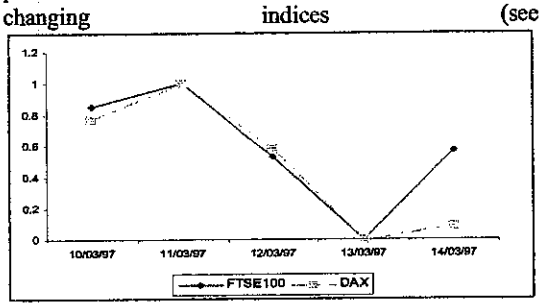
III. **Document Display:** This comprises two parts:

   a. **Document Titles:** Clicking a dot (date) on any of the time series, displays the corresponding date's news titles.
   b. **Document Content:** The content of any document title can be viewed by clicking that news title.

IV. **Document Analysis:** Whenever a document title is selected from the news list, the extracted sentiment keywords along with the frequencies are displayed in "Document Keywords" area. Positive sentiment keyword analysis details appear under the title of "Upward Movement Indicators" and negative sentiment keyword analysis details appear under the title of "Downward Movement Indicators".

4.3. Case Study

In order to investigate the effectiveness of the SATISFI system a *real-life* case study was identified for us by a capital management and research consultancy, JRC[1] Berlin. According to JRC market experts, there is a well-recorded turning point, from a low position on a day before to a high on the day followed by a low on the day after, in the European Stock Markets. Two of the major stock exchanges are the London Stock Exchange (with the FTSE100 index of 100 leading UK companies) and its German counterpart (the DAX 100 GDX). The high point of FTSE100 (and DAX 100) in the recorded turning point case study is that of the 11th March 1997 and two low points are on the 10th and 12th March 1997: a week of changing indices (see



[1] JRC is the lead partner in EU 5th Framework sponsored GIDA project.

Figure 8)

In order to see whether there was a corresponding change in the sentiment related to the UK markets to the FTSE100's behaviour, we selected a week's financial news from Reuters (10th to 14th March 1997) and extracted the potential sentiment expressing words according to our heuristic (see Section 3). There were 15,387 news items (45 MB) produced by Reuters News during that week. Figure 9 shows the correlation between the FTSE100 and the upward movement indicator: the Pearson correlation moment is 0.88; visually the change in FTSE100 and the upward movement indicator appears to be almost identical. We have also plotted the downward movement indicator, which is anti correlated with both the FTSE100 and the upward movement indicator (correlation moments of -0.85 and –0.57 respectively).

Two major feedbacks were received from JRC. One feedback was to see the positive news when 'positive sentiment time series' was clicked and to show negative news when 'negative sentiment time series' was clicked. The other feedback was to show the major news that caused the turning points.

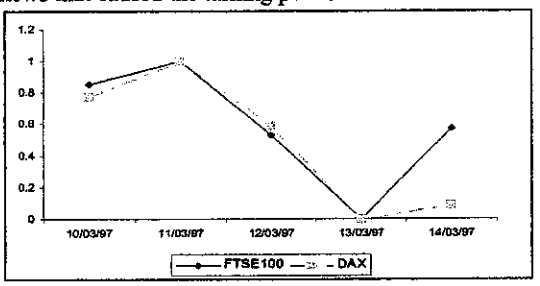


Figure 8. The chart displays normalised DAX and FTSE100 index between 10-14.03.2002.

According to JRC: Rather than displaying news headlines, visualisation of the sentiment of the news is much more beneficial in terms of providing a general overview of the news with respect to stock market in a quick way.

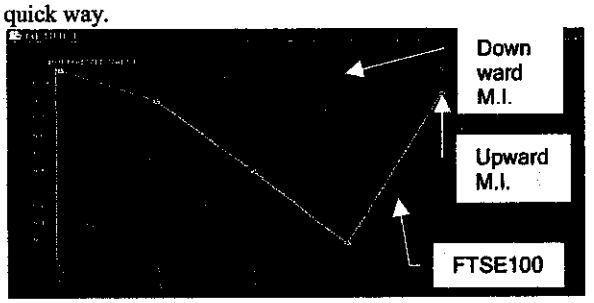


Figure 9. The SATISFI interface depicts three time series together: FTSE index (blue), upward movement indicators series (green) and downward movement indicators (red) series.

5. Afterword

System SATISFI shows how it is possible to visualise the data and information contents in two different modalities. The system is currently under evaluation by JRC Berlin and by the reviewers of the GIDA project appointed by the EU's Fifth Framework Programme for Information Sciences and Technologies. The initial results are interesting. The end-users have suggested that instead of using SATISFI as a prediction system, it should be used in investigating past behaviour of the market. For instance, using the visualisation system to find out which of the turning points in the value of an instrument (from *down to up* or *up to down*) were caused by a specific news item.

SATISFI provides a simple visualisation technique for identifying and displaying a complex feature of the news it receives. Furthermore, its ability to receive and display a numerical time series of financial data, and its ability to correlate the series with an index of market sentiment should be its notable feature.

6. References:

[1] C.Chen, *Information Visualisation and Virtual Environments*, Springer Verlag, London, 1999.

[2] J.Saltz, J.Steinbach. "Wayfinding in Layered Information Worlds", *CODATA Euro-American Workshop Visualization of Information and Data: Where We Are and Where Do We Go From Here?*, 24-25 June 1997, pp. 6-7.

[3] T.N. Bulkowski, *Encyclopedia of Chart Patterns*, John Wiley&Sons, New York, 2000.

[4] Reuters NewsML Showcase: http://about.reuters.com/newsml (accessed 30.04.2003)

[5] Investorwords Financial Glossary: http://www.investorwords.com (accessed 30.04.2003)

[6] Yahoo Finance: http://finance.yahoo.com (accessed 30.04.2003)

[7] Prophet.Net.: http://www.prophet.net (accessed 30.04.2003)

[8] Financial Times: http://www.ft.com (accessed 30.04.2003)

[9] A.Inselberg, "Multidimensional Detective", *Proc. of IEEE Information Visualization '97*, Phoenix, AZ, 1997, pp. 100-107.

[10] Conklin, N., Prabhakar, S., and North, C., "Breakdown Visualization: Multiple Foci Polyarchies of Tuples and Attributes", *Proceedings of ACM CHI*, March 2002.

[11] Teletrader: http://www.teletrader.com (accessed 30.04.2003)

[12] Smartmoney: http://www.smartmoney.com/ (accessed 30.04.2003)

[13] Neovision Tradermaps: http://www.neovision.com/solutions/tradermaps/ (accessed 30.04.2003)

[14] K. Ahmad, D. Cheng, T. Taskaya, S. Ahmad, L. Gillam, P. Manomaisupat, H. Traboulsi, A. Hippisley, "The mood of the (financial) markets: In a corpus of words and of pictures", *Proceedings of Corpus Linguistic 2003*. Lancaster University, UK
http://www.computing.surrey.ac.uk/iemmg/publications.html

gs of the Seventh International Conference on Information Visualization (IV'03)
/03 $17.00 © 2003 IEEE

Summarization of Multimodal Information[1]

Saif Ahmad, Paulo C F de Oliveira and Khurshid Ahmad

Department of Computing, University of Surrey
Guildford, Surrey, GU2 7XH, UK
{s.ahmad, p.oliveira and k.ahmad}@surrey.ac.uk

Abstract

Information Summarization is one of the key challenges for current and future information systems. In this paper, we will outline a system that comprises modules for summarizing texts and time series to study the link between the two. Summaries of texts are generated using a lexical analysis of cohesion in texts focusing on key sentences that provide cohesion: by implication, these are the sentences that comprise chief points of a given text. Time series summarization is accomplished using the so-called wavelet analysis to separate out the trend, cyclical fluctuations and autocorrelational effects and generating verbal signals to describe each phenomenon. Finally, we present a case study performed on the UK financial market with regards to multimodal information processing, namely textual and numerical summarization.

1 Introduction

The Oxford English Dictionary defines multimodal as "characterized by several different modes of occurrence or activity; incorporating or utilizing several different methods or systems". With the advent of the Internet and online data vendors (for example Reuters and Bloomberg), data availability is no longer a problem. The information provided by these data vendors is not restricted to text alone. In fact, information is provided in different modes ranging from text, graphics, video, sound, etc. If we look at sectors of a financial market as reported in financial newspapers, we find information with two distinct modes of expression namely textual and numeric. The textual information comprises financial news stories whereas the numeric information is essentially time series plots of financial instruments. This widening concept of "information or knowledge" makes it necessary to process multimodal information with a unified perspective, which demands for integrating various techniques, such as image and video processing, time series analysis and general text processing techniques.

Typically, summaries are understood as abridgements of natural language text documents – chief points, sum or substance of a matter (Mani, 2000). More recently, the notion of finding "chief points" has been extended to time series (Boyd, 1998) and (Sripada et al, 2001), and to video sequences. Perhaps sometimes in the future, one can talk about a system which will be able to synthesize the chief points about a matter from texts, and if relevant, time serial data and video sequences.

In this paper, we will outline a system that comprises modules for summarizing texts and for summarizing time series. Summaries of texts are generated using a lexical analysis of cohesion in texts focusing on key sentences that provide cohesion: by implication, these are the sentences that comprise chief points of a given text (Hoey, 1991). Time series summarization can be accomplished using the so-called wavelet analysis (Daubechies, 1996) to separate out the trend, cyclical fluctuations and autocorrelational effects and generating verbal signals to describe each phenomenon. The text and time-series components are used in the analysis of financial market movements in conjunction with the SATISFI system (Taskaya and Ahmad, 2003) developed at the University of Surrey under the auspices of the EU sponsored GIDA project (IST-2000-31123). Traders want to make a note of "turning" points in the time series data, i.e. when did the market start to fall or rise or what were the days on which the market fell or rose. Subsequent to this identification of tuning points, they wish to check on the news prior-to, on, or after-the rise or fall. Here, accurate summaries of financial news texts appear to be useful in that a mere display of a headline seldom suffices, as headlines do not have the discriminating power for examining the content of the texts.

In financial newspapers we often come across a news story that talks about the performance of the market and shows a time series as a "picture illustration" accompanying the text. Figure 1 shows an excerpt from the *Financial Times Online* (www.ft.com). The picture illustration shows the FTSE 100 time series while the accompanying text talks about the markets in London. The graph of the FTSE 100 shows an *uptrend* in the latter half and a text defines this *uptrend* in percentage and even gives the value to which it rises (highlighted yellow in Figure 1).

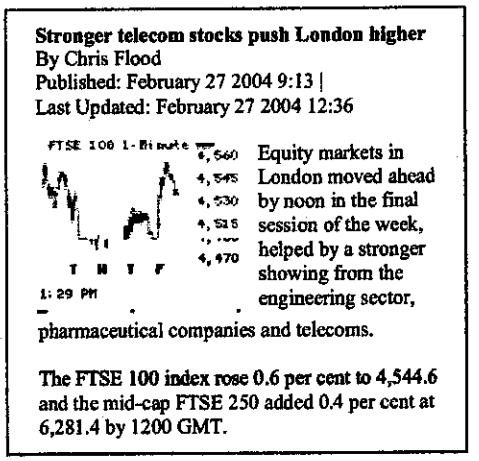


Figure 1: Excerpt from *Financial Times Online*

Here, the concise reporting about the London markets can be regarded as a textual summary whereas the description of the FTSE 100 trend (highlighted yellow) can be regarded as a time series summary.

1.1 Chief Points of Collateral Texts

Consider again the news accompanying the time series in Figure 1. The news headline is one of the main chief points: 'Stronger telecom stocks push London higher'.

This is a summary, which, however, requires considerable real-world knowledge: London stands for the London Stock Exchange and 'higher' is an abbreviation for 'higher values of the index'. The summary appears quite cryptic without the background knowledge of the

[1] To appear in the proceedings of LREC 2004, Lisbon, Portugal.

semantic relations that may exist between the keywords in the headlines.

The headline of the news item introduces the topics to be discussed, the so-called topic opening sentences; the first sentence in the body has some new words (engineering, equity market, pharmaceutical) but is dominated by words (or topics) already introduced – the so-called central sentences. The text in the Figure 1 is short, otherwise in longer texts there are sentences that introduce no new topics – all words have been introduced in the topic opening and central sentences.

Some of the sentences have no links (the marginal sentences) and others have above average of number of links. The sentences with above average links can be collated and be used as a surrogate for the original document – a summary.

This summary is more informative than, say, the headline alone. The cohesion link-based network shows the repeating textual patterns in a manner analogous to the time series with its cycles, fluctuations and trends – all the result of the distribution of frequently and rarely used patterns of number distribution.

1.2 Chief Points of a Time Series

The time series in Figure 1 can be summarized by suggesting that it has three major peaks and two major valleys and additionally has a number of minor peaks and valleys. The average value of the FTSE index is around 4500 and there is an observable *uptrend* Monday onwards. Furthermore, one can argue that the time between two peaks – one on Tuesday and the other on Friday is 72 hours and perhaps this cycle will persist. However, there are sections of rapid fluctuations – during Tuesday and Thursday.

The "chief points" of a time series are thus the cycles, trends and fluctuations. Violently fluctuating time series, when the first difference varies between positive (a rising series), or negative (a falling series) are said to have high *volatility*. If all the predictable patterns including cycles, trends and fluctuations are subtracted from a time series, then what is left is a set of random numbers – the so-called "shocks" in the system. These numbers are usually distributed according to a random distribution function with well-defined movements. However, if different random distributions are required over a period of time then the series is called *nonstationary*.

A time series is defined usually as an ordered series of numbers. The various changes in the value of the series have different return periods – cyclical changes have different periods, fluctuations or more accurately volatility patterns emerge over a period of time and then never recur and different volatility patterns appear at different times. These return periods manifest themselves as the so-called turning points or edges in a time series.

2 The Approach

Figure 2 outlines the basic approach we adopt to process multimodal information. We analyze data provided by Reuters, comprising 5000 stories per day (c.200-500 words/story) in over 150 ad-hoc topics, with a volatile time series – tick data from currency trading comprising 50 ticks per/second over an 8-hour period daily.

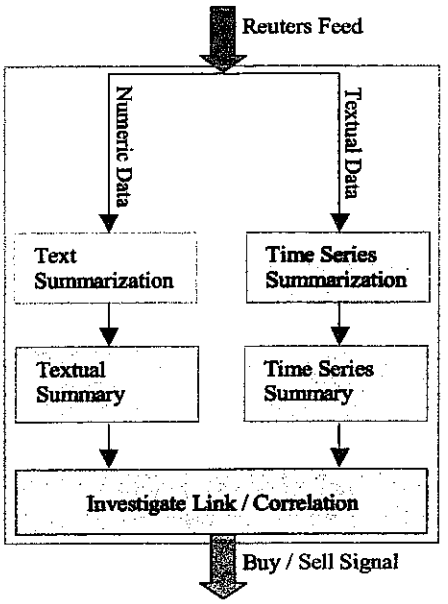


Figure 2: Synthesis of multimodal information

The traders mine and fuse this semi-structured and volatile data visually to generate buy/sell signals. Our framework (a) **summarizes** financial news articles using linguistic resources (e.g. lexical cohesion); (b) **summarizes** the locally-volatile financial time series in terms of the turning points using wavelet analysis; and (c) **facilitates** the visualization of the buy/sell signal. In this section, we will define our text summarization and time series summarization modules in greater detail.

(a) Text Summarization

This module is a computer implementation based on a linguistic theory of text organization called lexical cohesion (Hoey, 1991). Lexical cohesion is the tendency of the sentences in a text to carry information about a certain topic through related words and that provides quality of unity to the text. Hoey has argued that there is a very important connotation of marking cohesion explicitly in a text, which is lexical repetition. In fact, this approach uses the most frequent words of the text, their variants and conceptual relationships to establish connections between sentences in the text. The sentences that have a strong degree of association are selected for the summary. Two key notions were proposed by Hoey. The first key notion is *links* – which occur whenever there is a repetition of an item in two separate sentences. The second notion is *bond* – which is established whenever there is an above-average degree of linkage between two sentences. He stressed that the number of *links*, which constitute a *bond*, is relative to the type of text and to the average number of *links* in the text, but normally, three *links* constitute a *bond*, in order to avoid accidental repetition. Figure 3 is an excerpt from a financial news file collected from Reuters' Website (www.reuters.com) and illustrates both concepts. For example, the two sentences (nos. 18 and 20) in Figure 3 are bonded by four links, which together constitute a bond.

18. In other news, Hewlett-Packard said preliminary estimates showed *shareholders* had approved its purchase of *Compaq Computer* -- a result unconfirmed by officials.

20. In a related , *Compaq shareholders* are expected on Wednesday to back the deal, catapulting HP into contention against International Business Machines for the title of No. 1 *computer* company.

Figure 3: Example of bonded sentences

(b) Time Series Summarization

Traditional time series analysis methods involve decomposing a series into trend, seasonal variation, other cyclical changes and the remaining "irregular" fluctuations. However, this decomposition can be unique and effective only if certain assumptions about the underlying phenomena of the process are made, for example nonstationarity and volatility. Recently, it has been claimed that the so-called wavelet filtering provides insight into the dynamics of financial time series beyond that of current classical statistical methodology. A number of concepts, for example, nonstationarity, multiresolution, and approximate decorrelation have emerged from wavelet filters (Gençay et al, 2002). Wavelet filtering provides a natural platform to deal with the time-varying characteristics of real-world time series and is not restrained by the assumption of stochastic stability.

Wavelet analysis decomposes a time series into several sub-series (A1, D1, D2, ... DN), which may be associated with particular time scales. The interpretation of features in complex financial time series is made easy by first applying the wavelet transform and subsequently interpreting each individual sub-series. The extrema of the DWT sub-series detect major edges in the signal and hence correspond to most of the turning points. The recursive DWT filtering process removes long and short-term fluctuations from the signal in each recursion to give a linear trend in the end. Each recursive component is a time series in its own right. The major cycles within the various components (A1, D1, D2, ... DN) can be found using FFT techniques: this technique gives the frequency and amplitude for n/2 periods, where n is the number of data points in the time series. Different frequencies have different amplitudes – or loosely strengths – and in order to judge which of the n/2 cycles within the various components are important, the strength of the period of the signal can usually be determined by computing the Fourier coefficients. The variance change location is detected using the DWT on the volatility time series and making use of the so-called normalized cumulative sum of squares (NCSS) index.

3 A Case Study

Here, we report about a case study performed on the UK financial market for the month of May in 2002. The textual dataset comprises six financial news stories per day in May 2002 whereas the numeric data is the FTSE 100 index for the same period.

Figure 4 below shows the FTSE 100 time series for May 2002 with the trend (green line) and three most significant turning points (circled red) identified by our system.

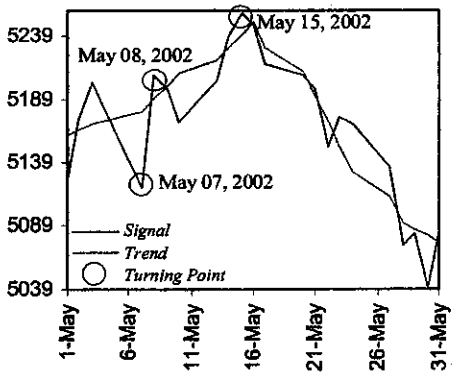


Figure 4: FTSE 100 (daily) for May-2002 showing turning points and trend identified by system

Table 1 shows our system's summary output (verbal signal generation) for other features extracted from the FTSE 100 time series data for May 2002. By looking at the system output for the FTSE 100 (Figure 4 and Table 1) we conclude that May 15, 2002 signifies a major turning point in that the trend is an uptrend until this point and a downtrend thereafter.

Feature	FTSE 100 (May 2002)
Trend	Uptrend from 1st to 10th day $S_L'(t) = +(8.89)\,t + 5151.79$
	Downtrend from 11th to 22nd day $S_L'(t) = -(15.99)\,t + 5426.98$
Dominant Cycle	2.66 days Starts on the 2nd day
Variance Change Occurs on the	5th day

Table 1: FTSE 100 summary output for May-2002

As discussed earlier, traders often want to know the news on and around financial turning points to make buy/sell decisions. In such a scenario, text summarization of news articles published on the day of the turning point and around it gains considerable importance. Table 2 shows the output from our text summarization module for the May 15 turning point, one day before it and one day after it. We have set the text summarization program to produce a one-sentence summary for each of these days so that we can correlate the news easily with the FTSE 100 time series. In Table 2, May 14 summary is good in the sense that it talks about the FTSE 100 "opening higher" the next day. This is supported well by the FTSE 100 index, which shows a peak the next day (May 15) in Figure 4. However, the time series module defines May 15 as a turning point and hence the index or the market is expected to fall from this point onwards.

Day	One-sentence News Summary
May 14	Financial bookmakers in London expect the FTSE 100 to open 10-16 points higher after a close of 5,204.8 and France's CAC-40 to start two points stronger than its close at 4,375.39 points.
May 15	Retailer Kingfisher dipped 3.5 percent to 372 pence after it made a two billion pounds cash call to help it fully buy French-based DIY group Castorama Dubois.
May 16	The firm reported pre-tax profits of 1.273 billion pounds in its year to end March, slightly above analysts' forecast range of 900 million to 1.25 billion pounds but down 28 percent on the year-ago period reflecting the new, slimmed-down BT Group.

Table 2: News on and around the May 15 turning point

The news itself on May 15 (the day of the occurrence of the turning point) in Table 2 talks about the retailer Kingfisher "dipping" – sort of indicating that the market could go down. The trend of FTSE 100 from this day onwards confirms it as a major turning point and the index continues to fall for the remainder of the month. Referring to Table 2 again, the news on May 16 (one day after the turning point) also talks about the British Telecom as "slimmed-down".

4 System SATISFI

The text and time series modules described herein are being used in conjunction with the SATISFI system developed at the University of Surrey (Taskaya and Ahmad, 2003). SATISFI is a multimodal information visualization and synthesis tool. Other capabilities of SATIFI include extracting market sentiment and events from news texts and then correlating it with the FTSE 100 time series; and news categorization using self-organizing maps. Figure 5 shows the SATISFI screenshot.

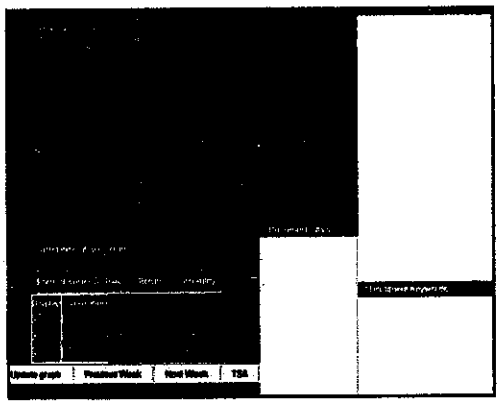


Figure 5: System SATISFI screenshot

The blue line in Figure 5 depicts the FTSE 100 index whereas the green and red lines depict the *upward* and *downward* market movement indicator time series. The user can correlate the market sentiment with the FTSE 100 index and can even shift the series forward and backward to check whether the news precedes the market movement or vice-versa. Moreover, each point on these time series can be double clicked to view the news headlines for that day. Furthermore these news headlines can be double clicked to display the entire news or its summary.

5 Afterword

Information about an event can be disseminated in a number of different modalities: science and technology, finance and business, linguistic modality and that of numbers, are amongst the key modes of interaction. The information supplied in these modes is voluminous and it is important to extract the chief points especially those points that show major changes related to the events. We have described a method for extracting turning points in a fluctuating and volatile time series and suggested ways in which this modality of numbers can be summarised in linguistic statements. Furthermore, we have attempted to introduce two possible uses of linguistic information that accompanies or complements the numerical information. First is the verificational use of the chief points of news streams for investigating the causes of potential turning points identified in the time series analysis. Second is the predictive use of news stories in predicting turning points based entirely on the class of news reported at a given time.

The method and analysis described above has been evaluated successfully but with a limited range of experts to both assess the efficacy of the time series summarisation and that of news summarisation. An extended trial involving a larger number of experts is required and is being planned. The other important development related to the trials is the release of the software on the web so the public at large can try. The method will be introduced in our work on the financial Grid project sponsored by the UK ESRC's initiative on eScience.

Acknowledgments

The authors wish to acknowledge the support of the EU-IST GIDA project (IST-2000-31123) and the UK ESRC's eScience project FINGRID.

References

- Boyd, S., "TREND: A System for Generating Intelligent Description of Time-Series Data", Proceedings of the IEEE ICIPS-1998.
- Daubechies, I., "Where do wavelets come from? – A personal point of view", Proceedings of the IEEE Special Issue on Wavelets, 84 (no. 4), pp. 510–513, April 1996.
- Gençay, R., Selçuk, F., Whitcher, B., "An Introduction to Wavelets and Other Filtering Methods in Finance and Economics", Academic press, 2002.
- Hoey, M., Patterns of Lexis in Text, Oxford University Press, 1991.
- Mani, I., Automatic Summarization, Amsterdam and Philadelphia: John Benjamins Publishing Company, 2000.
- Sripada, S., Reiter, E., Hunter, J., and Yu, J., "Modelling the task of Summarising Time Series Data using KA Techniques", In: Macintosh, A., Moulton, M., and Preece, A. (ed), Applications and Innovations in Intelligent Systems IX, Proceedings of ES 2001.
- Taskaya, T., and Ahmad, K., "Bimodal Visualisation: A Financial Trading Case Study", IEEE IV03, 7th International Conference on Information Visualisation, London, UK, 2003.

:UNDAMENTAL DATA TO SATISFI 'HE CHARTIST

/ Khurshid Ahmad together with Tugba Taskaya Temizel, David Cheng, Pensiri Manomaisupat, aif Ahmad, Lee Gillam, Haitham Trablousi and Matthew Casey

an market sentiment be visual-ed in the same way as share ices or any other financial set? If so, technical analysts ould use such charts for funda-ental confirmation of their chnical views. A team of searchers at the University of rrey present a prototype system hich they have developed for an U-sponsored project - GIDA Generic Information-based ecision Assistant) - the aims of hich include the development of method for automatically gener-ing trading signals from finan-al texts.

The financial trader receives market information in at least two forms: numbers d text. While various types of chart can be ed to display numerical data in a form that easy to digest and analyse, the financial ader is expected to read literally hundreds stories. Even if this only means reading e headlines or skim-reading, this is still a onumental task. Surveys of financial aders show that rather than using informa-e, complex techniques, simple visualisation chniques were always preferred (Saltz and einbach, 1997). Clearly, a system that can nultaneously deal with numeric and textual ta would be of benefit.

is the prototype of such a system being veloped for an EU sponsored project IDA) that we describe briefly. Despite the ct that textual information is not as dis-etely time indexed, this prototype demon-rates how news reports can be organised d analysed in order to assess and quantify e general mood of the trading market. The sessment is about the sentiment, which is nveyed in an individual news item: does e news item express a positive sentiment or a negative sentiment? Such expression is deeply embedded in language. Nevertheless, one can argue that the occurrence of words related to progress, profits, rising share prices may boost traders' confidence and the opposite, traumatic accidents, losses, falling share prices, will dent the traders' confidence.

The prototype:

- can manage a collection of news stories and historical time-series;
- can analyse the news stories for market sentiment;
- can correlate financial time-series with a time-series of keywords describing the positive and negative mood of the market.

Visualising **market sentiment**

Reports about the financial markets sometimes use metaphorical terms to indicate how the traders feel about the market: terms like 'bear' and 'bull' markets are well-known. News reports contains other expressions like: FTSE set to fall, Royal and Sun hit by asbestos suit, Energy tax hurts business.

Each of the underlined verbs fall, hit, hurts conveys a negative sentiment. On the other hand, verbs potentially like rise, jump, and climb may convey a positive sentiment. It is true that a word in natural language may have a range of meanings: the token rose may mean the flower rose; it can be a name, as in Mr/Ms Rose in addition to meaning 'increase in the value of a share or currency'. However, in financial texts tokens like rose, fall have a restricted meaning [14]. This is particularly true when these words are used in the following contexts:

{Shares, currencies} rose/fell {by} X percent
{Shares, currency} rose/fell {to} X percent

One problem-solving technique (heuristic) would be to use the frequency of these two verbs only if they occur in the context described above. The use of this heuristic allows us to quantify the market sentiment in

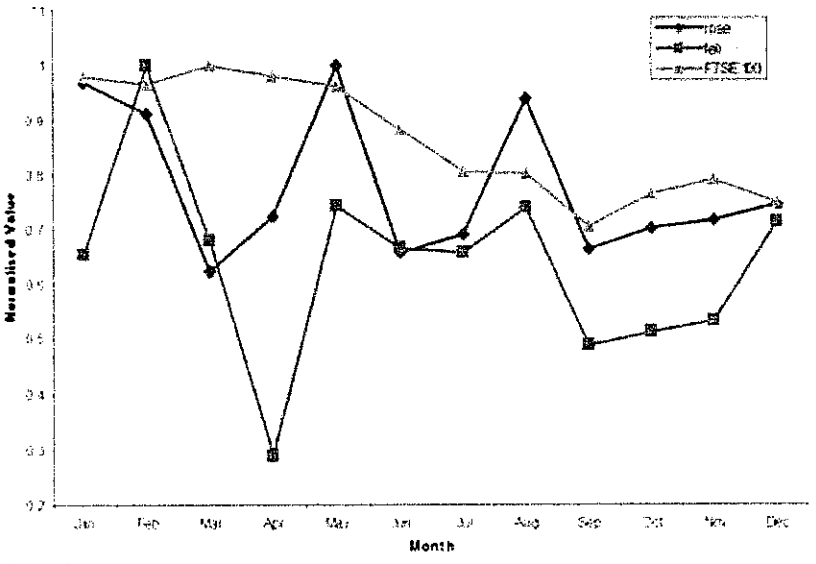
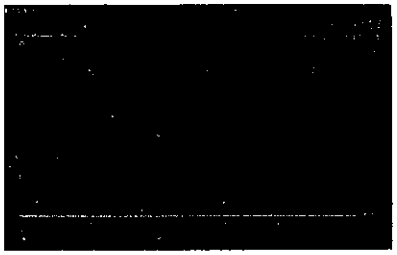


Figure 1.



gure 2. The visualisation system on the :b shows weekly news. The system dis-ays content of individual news items gether with the keywords associated ith the content.

much as a computer system can ever eval-.te any sentiment. We have analysed 2 mil-)n words of text comprising financial news pplied by Reuters for the year 2000. These xts were analysed by our text analysis sys-m, System Quirk (which performs text alysis for our prototype) and we extracted e frequency of the verbs rose and fell. iese frequencies were then plotted together .th the FTSE 100 index. Figure 1 shows)w the sentiment time-series correlates with TSE.

e have performed the analysis over a larger illection of financial news from Reuters er a 3-year period (2000-2002), comprising er 10 million words, and found that the uristics mentioned above hold as well. In ldition, many other sentiment verbs and epositions (up, down, adrift) have been entified and used to compute the market ntiment.

pically, a news agency like Reuters or oomberg will supply over 2000 stories per y comprising between 300-800 words. euters news stories give, amongst other formation, industry sector and country formation. Queries against a database of ws stories can be handled provided the ories have been pre-indexed on to the date publication, keywords characterising the ntent of the news, and, in our case, key rbs and prepositions characterising the sentiments, with positive and negative, expressed in and by each of news stories. Once a query has been successfully matched, it is important that the news story be displayed in full together with the frequency of the keyword, key verbs and prepositions. A visualisation system for viewing a collection of text, queried according to the various attributes of its constituent texts, has been developed in Java as a key component to our

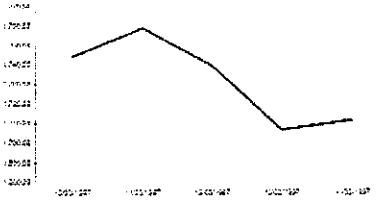


Figure 3. The visualisation system can work with Reuters news categorisation.

financial trading prototype (SATISFI: "Sentiment and Time-series: Financial Analysis System"). The system is written in Java and is accessible on the world-wide web (see Figure 2). Our visualisation system can work with Reuters country and industry sector categories together with Reuters supplied keywords (see Figure 3).

The SATISFI **system**

SATISFI was developed to show the correlation between a time-series of an instrument and market sentiment. The time-series of an instrument is seldom used in its raw form, rather derived forms are used: return and volatility. The index return is the logarithmic difference between two consecutive values and volatility is a measure of how fast or slowly the instrument moves up or down in terms of its value. Figure 4 shows how the prototype system works in terms of its four components: each helping to visualise either the market sentiment or an instrument's time variation. The value of the correlation helps the user to make a buy/sell decision.

SATISFI has four major components that have been integrated as shown in Figure 4.

1. Time-series Display: SATISFI can display three time-series at a time. These time-series comprise of FTSE-100 close index values, upward movement indicators and downward movement indicators. As discussed above, upward and downward movement indicators are the quantification of the market sentiment expressed in financial news. Over 70 terms each have been identified for conveying 'good' and 'bad' news. For example upward movement indicators would contain terms like 'up, rise, growth' etc. while down-

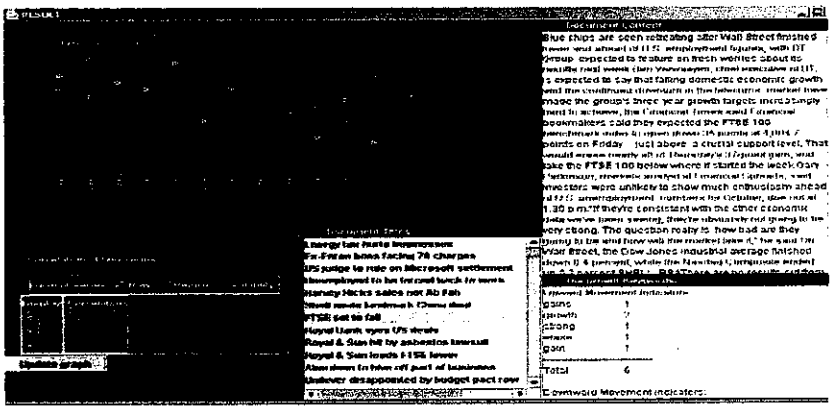


Figure 4. SATISFI prototype

ırd movement indicators would contain
:ms like 'down, fall' etc. The movement
dicator time-series are synthesized by
unting these movement indicator terms
thin the financial news published for a par-
ılar day. Each time-series is normalised for
oper display purposes. SATISFI is capable
displaying the above time-series in three
rms:
Raw form denotes the original time-series.
Return form refers to the logarithmic dif-
rence between two consecutive values.
Volatility (historical volatility) is the relative
te at which the time-series moves up or
)wn.

Time-series Correlation: SATISFI pro-
des the user the facility of cross correlating
'o series in any form (raw, return, volatility).
ny series can be shifted forward or back-
ırd and cross correlation recalculated to
:termine whether the market is followed by
e news or vice versa.

I. Document Display: This comprises of
'o parts:
Document Titles: Clicking a dot (date) on
y of the time-series, displays the corre-
)onding date's news titles.
Document Content: The content of any
)cument title can be viewed by clicking that
:ws title.

'. Document Analysis: Whenever a docu-
ent title is selected from the news list, the
:tracted sentiment keywords along with the
equencies are displayed in "Document
eywords" area. Positive sentiment keyword
ılysis details appear under the title of
Jpward Movement Indicators" and negative
ntiment keyword analysis details appear
ıder the title of "Downward Movement
ıdicators".

ase **Study**

. order to investigate the effectiveness of
e SATISFI system a real-life case study was
entified for us by a capital management
ıd research consultancy, JRC Berlin, (who
e also the lead partner in the EU spon-
sored GIDA project). According to JRC market experts, there is a well-recorded turning point, from a low position on a day before to a high on the day followed by a low on the day after, in the European Stock Markets. The two major stock exchanges are the London Stock Exchange (with the FTSE100 index of 100 leading UK companies) and its

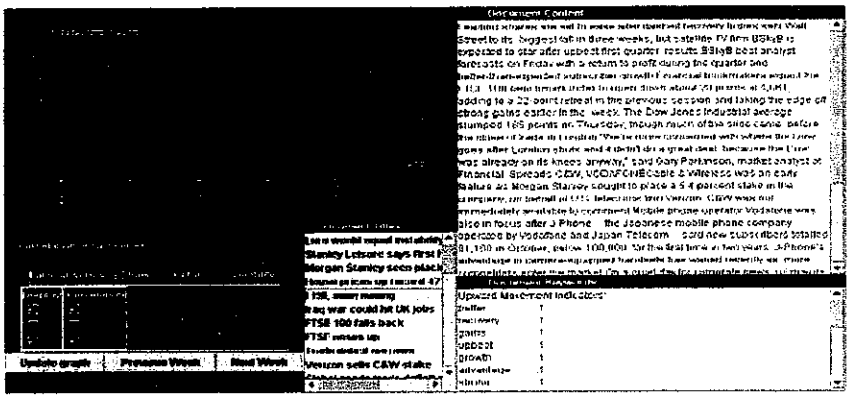


Figure 5. The chart displays DAX100 index between 10 and 14 March 2002.

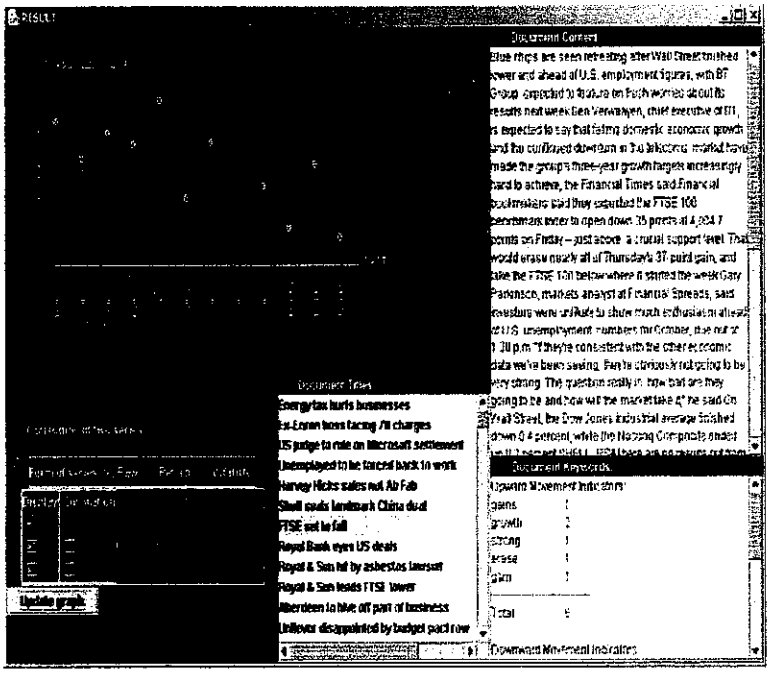


Figure 6. The SATISFI interface depicts three time-series together: FTSE index (blue), upward movement indicators series (green) and downward movement indicators (red) series.

"SURVEYS OF FINANCIAL TRADERS SHOW THAT RATHER THAN USING INFORMATIVE COMPLEX TECHNIQUES, SIMPLE VISUALISATION TECHNIQUES WERE ALWAYS PREFERRED."

ıen German counterpart (with the DAX 00 GDX). The high point of FTSE100 ınd DAX 100) in the recorded turning point ase study is that of the 11th March 1997 nd two low points are on the 10th and 12th larch 1997: a week of changing indices (see 'igure 5).

n order to see whether there was a corre-ponding change in the sentiment related to ıe UK markets to the FTSE100's behaviour, ve selected a week's financial news from euters (10th to 14th March 1997) and xtracted the potential sentiment expressing ords according to our heuristic. There were 5,387 news items produced by Reuters dur-ıg that week. Figure 6 shows the correlation etween the FTSE100 and the upward ovement indicator: the Pearson correlation ıoment is 0.88; visually the change in TSE100 and the upward movement indica-or appears to be almost identical. We have lso plotted the downward movement indica-or, which is anti correlated with both the TSE100 and the upward movement indica-or (correlation moments of -0.85 and -0.57 espectively).

Two major feedbacks were received from RC. One feedback was to show the positive ews when 'positive sentiment time-series' vas clicked and to show negative news when ıegative sentiment time-series' was clicked. The other feedback was to show the major ews that caused the turning points.

According to JRC: Rather than displaying news headlines, visualisation of the senti-ment of the news is much more beneficial in terms of providing a general overview of the news with respect to stock market in a quick way.

Afterword

SATISFI shows how it is possible to visualise data and information contents in two differ-ent modalities. The system is currently under evaluation by JRC Berlin and by the review-ers of the GIDA project appointed by the EU. The initial results are interesting. The end-users have suggested that instead of using SATISFI as a prediction system, it should be used in investigating past behav-iour of the market. For instance, using the visualisation system to find out which of the turning points in the value of an instrument (from down to up or up to down) were caused by a specific news item. SATISFI provides a simple visualisation technique for identifying and displaying a complex feature of the news it receives. Furthermore, its abil-ity to receive and display a numerical time-series of financial data, and its ability to cor-relate the series with an index of market sen-timent should be its notable feature.

Khurshid Ahmad is Professor of Artificial Intelligence at the University of Surrey and Principal Investigator of the GIDA Project. Taskaya Temizel and Cheng were Research Assistants of Prof Ahmad on the Project and worked on visualiza-tion and document manager respectively. Manomaisupat, S.Ahmad, and Trablousi are docotoral students of Prof Ahmad and researched on text categorization, time series analysis and local grammar respectively. Gillam is a Research Assistant of Prof Ahmad and has worked as an information architect on a number of EU projects. Casey was a Project Co-ordinator and now Lecturer in Neural Computing. All the authors are based at the Department of Computing, University of Surrey.

References:

J.Saltz, J.Steinbach. "Wayfinding in Layered Information Worlds", CODATA Euro-American Workshop Visualization of Information and Data: Where We Are and Where Do We Go From Here?, 24-25 June 1997, pp. 6-7.

K. Ahmad, D. Cheng, T. Taskaya, S. Ahmad, L. Gillam, P. Manomaisupat, H. Traboulsi, A. Hippisley, "The mood of the (financial) markets: In a corpus of words and of pictures", Proceedings of Corpus Linguistic 2003. Lancaster University, UK. http://www.computing.surrey.ac .uk/iemmg/publications.html