TrendFocus: Visualization of Trends in Financial News with Indicator Sets

Yik-Wai NG and Huamin Qu
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
ywng@ust.hk
huamin@cse.ust.hk

***Abstract*— Although many current visualization approaches to time-series data can show a good general trend of a topic or theme over time, most cannot help people gain a deeper understanding of key issues. They may illustrate how some key issues come up and disappear from time to time, but they typically provide limited information about relationships of interest. Financial and economic analysts are often interested in how financial or economic performance indicators interplay or change over time. With a set of well-defined and commonly used equations or indicators in the domain of Economics and Finance, we can refine our visualization by keeping track of these variables. In this paper, we make use of a line plot layout to track some of these indicators and leverage word clouds to reveal how these indicators are affected by the changes in social and economic environments.**

***Keywords*—Financial News Visualization, Trend Focus, Domain Knowledge Visualization .**

I. INTRODUCTION

In this era of big data, huge amounts of data or information are generated every day. Visualizing such a huge amount of information has become a challenging task for researchers, especially when dealing with time-varying data. However, we inevitably encounter a myriad of time-series data in our daily lives, like newspapers, stock information, social network data and the like. Analysts often want to dig out trends or patterns via data mining to make decisions based on the past history. Currently, different kinds of tools have been developed by different researchers for visualizing time-varying data, for example, SpiralClock [1] and Spiral Graph [2] for periodic data, Trendalyzer's [3] animation-based visualization for temporal patterns, TextWheel [4] and ThemeRiver™ [5] for thematic or topic evolution over time, GeoTime® [6] and Growth Ring Maps [7] for spatial temporal data, and so on.

In the domain of economics and finance, users are typically interested in trends or patterns of data over time. Although the current visualization tools for theme or topic evolution over time do a great job in tracking general trends, they may not tell us specific or domain related knowledge regarding the trends. For example, in the context of economics and finance, we may want to know how certain economic performance indicators interplay or change over time in the context of currently debated topics or themes. Most thematic visualization tools only give users a general idea of changes or trends of a topic, e.g. quantitative easing. But to know more about the impact and influence of a topic on an economy or financial market, users still have to look up lots of data by themselves to figure it out. To know more about the depth and meaning of a certain theme or topic, we rely on the current visualization methods, tailoring them to visualize specific financial and economic news.

In this paper, we incorporate the concepts of economic and financial equations in the design of visualization of financial and economic news. Since equation variables are well-defined by many economists, they can serve as good indicators for financial market performance over time. For example, there is some sort of implicate relationship between interest rate and inflation, and generally speaking, when the interest rate is low, inflation tends to be high. If we keep track of their changes together with changes in economic and financial topics, we can help users better understand market changes over time and hopefully make better investment decisions. We do this by analyzing semantic changes of those indicators at a sentence level. We similarly define how topic keywords correlate to certain indicators.

In summary, the major contributions of this paper are:

1. Integration of Wordle and line plots to keep track of the changes in both themes or topics and some economic and financial performance indicators.

2. Incorporation of economic equations for enhancing the visualization of some domain specific knowledge.

3. Verification of the implicit relationship among indicator variables by visualizing the given financial news data. Or if any abnormality happens, users can observe it and delve deeper into the details.

II. RELATED WORK

Much effort by different researchers has been dedicated to visualizing time series data to discover trends or thematic changes over time. The use of Wordle or Word Tag to aid the finding of keywords is very common.

A. Keyword Visualization

To find specific topics or themes over time, an algorithm to find their keywords is needed. We have different visualization tools to serve the purpose of visualizing the content of a given corpus and showing the keywords. Wordle [8] and Tag Cloud [9] are popular ones which visualize the significance of certain keywords by using different front sizes, colors, etc. Many document visualizations are based on the Vector Space Model

[10] to show document characteristics. Some count word frequency for extracting and ranking the keywords while others look for patterns and the main themes of documents. For example, tf-idf (term frequency–inverse document frequency) [11], a numerical statistic which reflects how important a word is to a document in a collection or corpus, is commonly used.

Figure 1 Visualization of economic and financial news with TrendFocus to gain more insight into some economic performance indicators as well as grasp overall trends of current issues

B. *Visualization of Trends or Changes Over Time*

For visualizing trends or thematic changes, many tools have been developed. Quite a number of them are based on information theory developed by Shannon [12]. Some analyze the semantic relations between the words, for example, co-occurrence relations. Others use entropy to estimate the relationship between information: mutual entropy is used to estimate information shared between two clouds of information; conditional entropy estimates a word cloud's significant values, etc.

Cui et al. [13] have developed a trend chart that summarizes content changes over time. It integrates a chart diagram and dynamic word clouds which are context preserving (stabilizing word placement across different word clouds). The words are placed according to their semantic relationships, and the trend chart is developed based on information theory, e.g. mutual entropy and conditional entropy. Relationships or trends can be further classified as macro or micro relationships. Cui et al. in their TextWheel tool [14] also tried to visualize large-scale news streams in the context of macro relations (between news articles), micro relations (between keywords) and macro-micro relations (between news articles and keywords).

To visualize the evolution of themes or trends over time, the metaphor of a river or current flow is most widely used. ThemeRiver by Havre et al. [15] is the pioneer in using the river metaphor to represent thematic variations over time within a large collection of documents. The flow from left to right serves as a timeline functionality. The change of the width of the "colored current" represents the strength variations of a topic over time. After that, more extended works enhanced the use of the river metaphor to represent thematic changes. These include Tag River [16] by Forbes et al., which considers the use of tag cloud in the river metaphor to better indicate the coherence of context between time segments.

C. *Semantics Mining*

Methods such as Speech and Syntactic Tagging, Rules Processing, Alias Expansion, Geographic Normalization, Named Entity Recognition, and Co-Occurrence Analysis have been used to aid accurate estimation of semantics. Other related work includes Wang et al.'s multi-document summarization [19], which is based on sentence-level semantic analysis.

III. OVERVIEW

In this section, we first state what research problems we are trying to address, and then we give an overview of how our system works to solve these problems.

A. *Research Problems*

We designed our system to answer some domain specific questions while we are revealing the general trends. In particular, we are answering the following questions:

1. What are the general trends, issues or topics evolving in the overall market over time? What are their impacts on economic and financial markets? Which variables (e.g. interest rate, inflation rate, etc.) of the market are they influencing in particular?

2. How do particular indicator variables of economic and financial markets change over time? Do they increase or decrease mildly or dramatically? How often are they mentioned or discussed, and how are they related to issues or topics currently debated, or what particular issues are they more related to?

B. *Our System*

The main components of our visualization system are shown in Figure 1: a trend chart for indicator variables, word cloud diagrams as the backdrop of the trend chart, and a control panel for convenient switching between different time intervals.

Time Interval Control Panel (at the bottom of Fig. 1): For temporal visualization, we must have a good control panel for traversal throughout the time domain. Inspired by the Google Finance [17] platform, we adopted a design which allows users to quickly traverse throughout the time periods. The yellow bar indicates the total length of the time interval that the system covers. Users can slide the two square gray boxes (left: start of the sub-interval; right: end of the sub-interval) to indicate a sub-interval to be drawn onto the main window. Users can lock the sub-interval length (by pressing the button in the left bottom corner) for traversal too. Quick selection of sub-intervals of 1 day, 3 days, 1 month, etc. is provided at the top of the window.

A Trend Chart (Fig 2): each line represents an indicator variable, and the color is for the convenience of classification only. The width of the line indicates how often the variable is mentioned in the news within a given date. If it is not mentioned, a dotted line is used to connect the gap. The dotted line in the middle represents the balance line, and if a point lies on it, the variable neither tends to increase nor decrease. If the variable is found to be above the dotted line, it means it tends to carry "increase" semantics. The further it is above, the relatively stronger is that meaning. It is also true for the opposite (under the dotted line). Note that the height of the line plot doesn't represent the absolute "semantic value" of a certain variable. When we click on certain node, detailed information (variable appearance frequency, net semantic value, related news articles and sentences) will be given upon inquiry (Fig. 3 Bottom).

A Word Cloud (Fig 1): this serves the purpose of providing more insights to the user. The font size of keywords represents their appearance frequency within the news documents during that given period. The color of the key is classified based on its correlation to the variables. If a keyword is more related to a particular variable, its color will be the same as the line color of that particular variable. When we click onto a certain keyword, detailed information (news articles and sentences in which the keyword and the indicator variable co-occur) will be given upon inquiry (Fig 3 Top).

IV. Design and Implementation

A. Trend Chart (Line Plot)

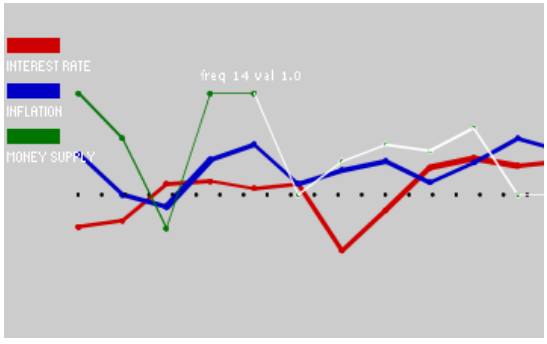

Figure 2. A trend chart showing the semantic values of the variables (Red: interest rate; Blue: inflation; Green: Money Supply). When the mouse moves over a node, the selected line will be highlighted and brief information such as the apperance frequency and semantic value will be given

1) Selection of Indicator Variables: We select the variables based on well-known and well-defined equations or relationships in the economics and finance field. Of course, we allow users to define their own set of variables too. The use of well-defined equations or relationships as the set of variables allows users to have a better and quicker understanding of the market. It also allows users to know more about how those variables interplay. If they have some domain knowledge regarding these equations or relationships, it also improves their visualization experience. For example, one set of variables $\{Interest\ Rate, Inflation, Money\ Supply\}$ (Fig. 2), is based on the Fisher Equation $1 + i = (1 + r)(1 + \pi)$, and one relationship (i.e., if money supply increases, interest rate generally decreases.).

Based on the equation and the relationship, we put the three variables together as a set. While doing the processing, we recognize that the variables may have different synonymous name forms. For instance, interest rate may be mentioned in other forms such as interest rates, federal rate, interbank borrowing rate, etc. We thus also count its amorphous or synonymous forms.

2) Visualization of Appearance Frequency of Indicator variables: This is represented by the stroke weight of the line segment between the two nodes of the curve (Fig. 2). The stroke weight is within the range of $[0.7,8]$. We take the appearance frequency of the indicator variable in the two time nodes, and get the average. Then, $LineStrokeWeight = \frac{AverageFreq}{MaxFreq} * StrokeRange$.

3) Calculation of the Semantic Value of Variables: We do "localized sentence level" semantic estimation. Since we are estimating the semantic value of the variables only, we localized the nearby words around the variable keywords in the sentence and perform the semantic calculation based on a list of "increase/decrease semantic" wordings. For each "increase/ decrease semantic" wording, we assign a weight (Table I). For example, for the word "skyrocket", we assign the highest weight.

TABLE I. INCREASE / DECREASE SEMANTIC

Increase Semantics	*Weight*	*Decrease Semantics*	*Weight*
high	1	low	1
increase	2	decrease	2
raise	2	fall	2
higher	2	lower	2
skyrocket	3	plunge	3
……	…	……	…

Example 1. The sentence: "The interbank interest rate has surged over the past two weeks, making it harder for companies, particularly smaller ones, to borrow money." Our variable keyword is "interest rate". By analyzing the nearby 4-5 words, we spot the word "surged", which signals a drastic increase. So, we add 3 to the count of *IncreSemanticReferences*

Then, to calculate the value of the semantics for a variable, we evaluate

$$EntityPolarity = \frac{IncreSemanticReferences - DecreSemanticReferences}{TotalSemanticReferences}$$

, which has a value between -1 and 1. If the value is 1, the variable is "increase dominant", which indicates an increasing trend.

B. Word Cloud

1) Keyword Generation and Layout: The keyword generation and layout algorithm is done by an open source program called WordCram [18]. It does the text analysis for extracting keywords and avoids collision during word placement.

2) Color of Keywords and Correlation to variables Calculation: In addition to leveraging the power of the existing word cloud generator (WordCram), we do more to analyze the relationship of each keyword with the variables. We determine which variables's color is used for the keyword color based on co-occurring relationships. We define $cooccurringValue = \frac{NumCooccuringSentence}{NumCooccuringDocument}$, where a co-occurring sentence is a sentence that contains both the variable and the word cloud keyword, and a co-occurring document is a document that contains both the variable and the word cloud keyword.

For example, if the $cooccurringValue$ is highest for the variable "interest rate", both the line plot of interest rate and that word cloud keyword will be the same color. If the value is zero for all variables in the set, the color of the word cloud keyword will be white, which indicates that this keyword occurs frequently but does not correlate to any variables in the set. It may imply that users should pay special attention to it. It may carry new or special information that users may be interested in or find useful.

Figure 3. Details Inquiry Window: a) Word Cloud Key Related Information (the top example): the window show you the details of documents and setences which contain both the word cloud keyword and the variables. The keyword is highted yellow, while the variable is hightlighted its assigned color. b) Variable Related Information (the bottom example): the window shows the detailed documents and sentences that contain the variable

3) Details Inquiry Interface (Fig 3): When users find something strange or surprising, they may want to know more about the details of the text. Thus, we provide users with the ability to view the list of sentences that contain the keywords or the indicator variables, which allow a quick understanding of the situations. If users want to go further to read more information, they can also view a list of related documents. In both the viewing of sentences and documents, users can type in additional keywords (according to their own interests) to be highlighted in the search bar for their reading convenience.

V. Case Study and Discussion

Recently, if we have paid attention to finance-related news or financial markets, we know that the US Federal Reserve has been considering tapering off Quantitative Easing (QE). This has become one of the most debated topics recently. There are different opinions on this subject and some uncertainty about the Fed's policy changes. Tapering off of QE seemed to be planned but the schedule and speed were not certain. Let's use our tool to explore and see if pertinent information can be visualized or stand out. In the case study, we pulled about 3000 newspapers from the ProQuest database, in which the publishers include but are not limited to The Financial Times, The Wall Street Journal, The Guardian, New York Times and China Today.

A. Message from our Trend Chart

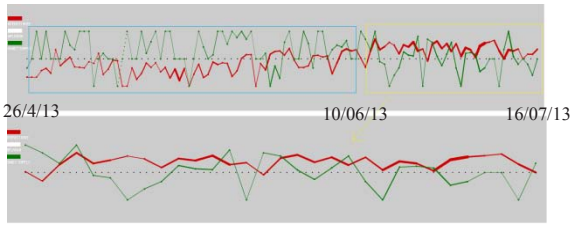

Figure 4. The Trend Chart from our case study

In our visualization tool, observing the period from 26th April 2013 to 16 July 2013, we have an idea of the general trend of the interest rate as well as money supply (Fig. 4 at the top). Around 10th June 2013, we can see a turning point in the interest rate and money supply. Since then, the interest rate has tended to be increasing, while the money supply has tended to be decreasing with some fluctuation (Fig. 4 at the bottom). These two curves give us an idea of how these two important economic performance indicators have been changing. Generally speaking, the two indicator variables also maintain their implicit relationship from the perspectives of economic theories, that is, interest rates should increase when money supply decreases. Next, we will look further into the details of the information in that time interval.

B. Information from Word Clouds

1) Find the Reasons: The information from word clouds very likely tells the reasons for these changes, and their impacts. The yellow circles highlight some likely reasons for the increase in the interest rate (Fig 5). It might be due to "Fed", "Reserve", "QE", "bond", "new", "cut", "tapering", "job", "recovery"..... Guessing from these keywords, we can

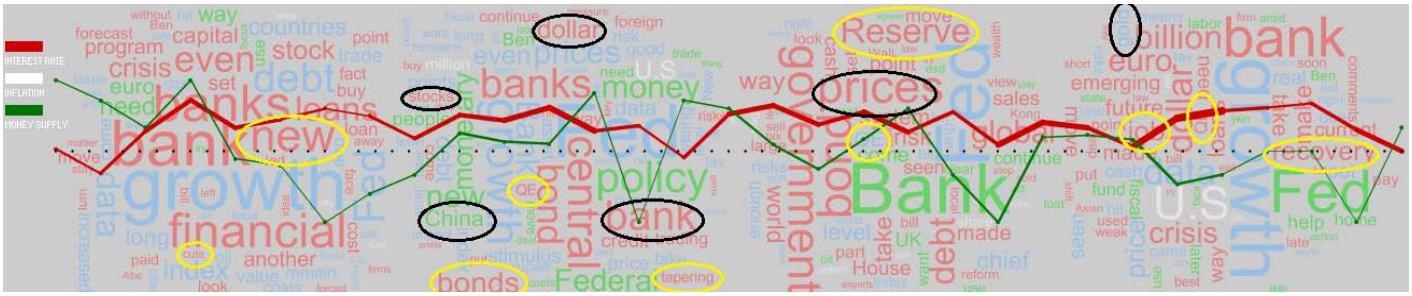

Figure 5 A Trend Chart with Word Clouds from our case study; yellow circles: some of the possible reasons for the changes of the indicator variables; black circles: some of the possible impacts of the changes.

surmise that the increase in interest rate is due to the Fed's cut in QE (bond buying). Some people might further believe this means an economic recovery. To further verify the ideas, we use the details inquiry window to look into sentences, or even documents. The list of sentences below provide a deeper interpretation:

"Long-term interest rates shot up after the Fed statements, and stock prices have tumbled."
"The Fed is also understating the impact of its tapering plan on interest rates."
"The Fed's tapering of QE 3 is also expected to be gradual and interest rates kept on hold until 2015."
"a rise in bond yields consequent to US recovery and expectations of interest rate increase."
"A reduction in quantitative easing QE leads to rising bond yields which together with the prospect of rising interest rates increases the opportunity cost of holding gold which pays no income."
.....

After looking into some of the sentences that contain both the highly frequent word cloud keys and the indictor variables, we know that people have different opinions on the QE plan, but generally people expect a cut and an increase in interest rates.

2) Find the Impacts: Similarly, we could spot some keywords which might be the impacts of the increase in interest rate (Fig. 5 black circles). Below is a list of informative sentences:

"A reduction in quantitative easing QE leads to rising bond yields which together with the prospect of rising interest rates increases the opportunity cost of holding gold which pays no income"
"The interest rates that banks charge to lend to each other shot up Thursday, and lending between banks nearly seized up after the People's Bank of China uncharacteristically failed to intervene to relieve a cash squeeze"
"Bank stocks have been hit amid worries about higher interest rates on their balance sheets"
"Long-term interest rates shot up after the Fed statements, and stock prices have tumbled"
"Falling bond prices mean higher long-term interest rates, and this has undermined equity markets, especially the equity prices of those firms that employ a large amount of debt financing."
"Rising interest rates and a strong US economy would be a shock supporting a stronger dollar and ending the search for yield that had driven so many assets from defensive shares through junk bonds to emerging-market local currency debt."
.....

By looking into these sentences, we can deduce from the information (assuming it's true) that the impacts are actually quite widespread: a rise in borrowing costs, a negative effect on housing, gold and some other real assets markets, a decrease in stock prices, a strengthening of the US dollar, etc.

3) Find the Topic Evolution: By observing and comparing the word clouds over time, we could get some hints about the emergence and evolution of an issue or topic. For example, the keyword "QE" (Quantitative Easing) appears in the second, third, and fourth word clouds with increasing size, showing that people probably started to talk about QE on 18th June 2013 and it became more and more popular since then. Clicking into the three instances of the keyword, and looking into details of sentences, we can see the evolution of the context too.

Also, the change in color reveals some sort of information regarding its topic content. The first and last QE instances are in red (more correlated to interest rate), while the second one is in blue (more correlated to inflation). We found out that people did mention the QE generally, exploring what it generally means and its most direct influences. Later on, people thought more broadly about other perspectives while exploring the topics.

C. Discussion

The case study shows that our tool is capable of good visualization of some significant events, such as Quantitative Easing (QE). It not only reveals some trends in the aspects of topic or content evolution, but also the semantic changes of some indicator variables that users are more interested in. However, there are still some shortcomings that need to be overcome in order to visualize the trends and semantics more accurately:

1. The analysis on the indicator variables is in global context. For some localized changes, the results might be distorted and not clear. For example, the interest rate in China may be rising while it may be decreasing worldwide.

2. The current semantic analysis method sometimes cannot accurately estimate the meaning. For example, when the "increase/decrease semantics" wordings are too far way, or the text is in double negation or other special writing styles, we cannot get a good estimation of the semantics, or the information may even be distorted. A more vigorous semantic analysis should be performed.

3. The correlation between a Wordle keyword and an indicator variable is roughly estimated based on the number of co-occurring sentences and documents. Maybe a more sophisticated method for analyzing the correlation should be performed.

4. The Wordle keywords are all just single words. If our system supported phrases, then the interpretation of information from word clouds would have improved.

5. Above all, we are just visualizing the information contained in mainstream financial news, but we cannot be sure at its veracity. The tool just helps people quickly interpret a huge amount of news data. We cannot guarantee that the information truly reflects the actual markets. Several points are worthwhile to note: First, markets may not be free and may be manipulated by governments or large corporate entities. Secondly, the mainstream financial media's representation of facts, trends, news and important events may be biased and not representative or reality.

VI. Conclusion and Future Work

In this paper, we have presented a visualization system for large-scale financial news streams. We have combined the use of a trend chart and word clouds to convey information. While discovering a change or evolution in topics, we keep track of the semantic meanings of some domain knowledge related indicator variables that we might be more interested in. The implicit meanings or relationships among these variables in a given set may also reveal valuable information to users.

In the future, we plan to improve the accuracy of the semantic analysis and thus improve the usefulness of the visualization tool. We may apply some techniques which are used in the natural language processing in our work, so as to understand the text more thoroughly. Also, we may extend our ideas and methods to visualization of information in other domains, such as politics, helping us understand political changes. While carrying out experiments in our case study, we found a potential use of our tool to test some sorts of relationships or correlations, or use the tools to find some "undiscovered" relationships or correlations. The tool may facilitate teaching in finance or economics subjects as well.

Acknowledgment

This research was sponsored and supported by Undergraduate Research Opportunities Program (UROP 1000) of the Hong Kong University of Science and Technology.

References

[1] Dragicevic, P. & Huot, S.: SpiraClock. A Continuous and Non-Intrusive Display for Upcoming Events. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), ACM, 2002.

[2] Weber, M., Alexa, M., and M··uller, W. (2001). Visualizing Time-Series on Spirals. In Proceedings of the IEEE Symposium on Information Visualization (InfoVis), pages 7–14

[3] Gapminder Foundation: Gapminder Trendalyzer: http://www.gapminder.org/world/

[4] Weiwei Cui, Huamin Qu, Hong Zhou, Wenbin Zhang, and Steve Skiena. 2012. Watch the Story Unfold with TextWheel: Visualization of Large-Scale News Streams. ACM Trans. Intell. Syst. Technol. 3, 2, Article 20 (February 2012), 17 pages.

[5] Havre, S.; Richland; Hetzler, B. ; Nowell, L.: ThemeRiver: visualizing theme changes over time, Information Visualization, 2000. InfoVis 2000. IEEE Symposium

[6] Kapler, T. & Wright, W.: GeoTime, Information Visualization. Information Visualization, Vol. 4, No. 2, 2005.

[7] Bak, P.; Mansmann, F.; Janetzko, H. & Keim, D. A.: Spatiotemporal Analysis of Sensor Logs using Growth Ring Maps. IEEE Transactions on Visualization and Computer Graphics, Vol. 15, No. 6, 2009.

[8] Fernanda B. Viegas, Martin Wattenberg, and Jonathan Feinberg. 2009. Participatory Visualization with Wordle. IEEE Transactions on Visualization and Computer Graphics 15, 6 (November 2009)

[9] Fernanda B. Viegas and Martin Wattenberg. 2008. TIMELINES: Tag clouds and the case for vernacular visualization. interactions 15, 4 (July 2008), 49-52.

[10] G. Salton, A. Wong, and C. S. Yang. 1975. A vector space model for automatic indexing. Commun. ACM 18, 11 (November 1975), 613-620.

[11] Fabrizio Sebastiani. 2002. Machine learning in automated text categorization. ACM Comput. Surv. 34, 1 (March 2002), 1-47.

[12] C. E. SHANNON, A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.

[13] Weiwei Cui,Yingcai Wu ; Shixia Liu ; Furu Wei ; Zhou, M.X. ; Huamin Qu: Context preserving dynamic word cloud visualization, Computer Graphics and Application, Volume:30 Issue:6

[14] Weiwei Cui, Huamin Qu, Hong Zhou, Wenbin Zhang, and Steve Skiena. 2012. Watch the Story Unfold with TextWheel: Visualization of Large-Scale News Streams. ACM Trans. Intell. Syst. Technol. 3, 2, Article 20 (February 2012), 17 pages.

[15] Havre, S. ; Battelle Pacific Northwest Lab., Richland, WA, USA ; Hetzler, B. ; Nowell, L.: ThemeRiver: visualizing theme changes over time, Information Visualization, 2000. InfoVis 2000. IEEE Symposium

[16] Forbes A G, Alper B, H O Llerer T, et al. Interactive folksonomic analytics with the tag river visulization [OL].[2012-12-31].

[17] Google Fiannce: www.google.com.hk/finance

[18] WordCram: http://wordcram.org/

[19] Dingding Wang, Tao Li, Shenghuo Zhu, and Chris Ding. 2008. Multi-document summarization via sentence level semantic analysis and symmetric matrix factorization, in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08). ACM