行政院國家科學委員會專題研究計畫 成果報告

資訊內容與市場反應：文字探勘之應用
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 98-2410-H-343-010-
執 行 期 間 ：98 年 08 月 01 日至 99 年 08 月 31 日
執 行 單 位 ：南華大學會計資訊學系

計 畫 主 持 人：洪嘉聲
共 同 主 持 人：歐進士、蔡志豐
計畫參與人員：碩士班研究生-兼任助理人員：孫淑芬

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 99 年 12 月 03 日

Ⅰ. Background and motivation

In capital market, information always impacts the market. The sources of information are widely diversified. The information disclosures can be provided voluntarily or enforced by the regulations. Both of the two kinds of disclosures will convey some information and generate the capital market consequences. Most of information disclosed, except the financial statement or reports, is text-formatted and unstructured. Prior studies extracted some specific information from the unstructured information and showed the impacts on the capital market. They left lots of information unexamined. In order to fill this gap, this study first convert the text-formatted information into digital and structured information and then take advantage of data mining techniques to extract some rules which can show the words may cause some impacts on capital market.

Prior researches examined the information content of disclosure and their economic consequences. They extracted some specific information from the disclosures. For example, Han and Tan (2007) studied the investor's reactions to management guidance forms. Libby, Tan and Hunton (2006) investigated the impacts of management's earnings guidance on analysts' earnings forecasts. These two papers just took advantage of small parts of disclosure-the management's forecast and the guidance form. Gu and Li (2007) study the disclosures of innovation strategy made by 180 high-tech firms from 1992 to 1994. They reviewed each press release of sample firms to subjectively determine its relevance for the firm's technological innovation, and hand-collect disclosures about the strategy of innovation. This kind of research method leaves lots of information unexamined and is very time consuming. Besides, the results are likely biased due to the subjective judgments.

This study focuses on the impacts of press release on market reaction. The content of press release is unstructured and suitable for text mining. Text mining is a semi-automated process of extracting knowledge from a large amount of unstructured data and known as intelligent text analysis, text data mining or knowledge discovery in text (Feldman and Dagan, 1995). Text mining applications fall into two areas: exploring the textual data for its content, and then using the information to improve the existing processes. Both of applications can be referred to as descriptive mining and predictive mining.

The U.S. Securities and Exchange Commission (frequently abbreviated SEC) is a federal agency which holds primary responsibility for enforcing the federal securities laws and regulating the securities

industry, the nation's stock and options exchanges, and other electronic securities markets in the United States. In addition to the 1934 Act that created it, the SEC enforces the Securities Act of 1933, the Trust Indenture Act of 1939, the Investment Company Act of 1940, the Investment Advisers Act of 1940, the Sarbanes-Oxley Act of 2002 and other statutes. The SEC was created by section 4 of the Securities Exchange Act of 1934 (now codified as 15 U.S.C. § 78d and commonly referred to as the 1934 Act). The news of SEC should be neutral for capital market and fit for data mining.

The first phase of this study involves converting the text-formatted news into digital and structural information. Without this step, we cannot use data mining techniques to explore the relationship between news content and market reactions. By taking advantage of data mining, the second phase of this study is to explore the news content and market reactions. The main contribution of this study is to find the words in news that may affect the capital market.

Ⅱ. Literature Review

The information disclosures can be provided voluntarily or enforced by the regulations. The voluntary disclosures include management guidance, analysts' presentations, conference calls, press releases, internet sites, and other corporate reports. The disclosures enforced by the regulations include financial statement, footnotes, management discussion and analysis, and other regulatory filings. Both of the two kinds of disclosures will convey some information and generate the capital market consequences. Financial reporting and disclosure are important for management to communicate firm performance and governance to outside investors. Healy and Palepu (2001) provide a framework for analyzing managers' reporting and disclosure decisions in capital markets setting. They review the demand for disclosure, the effectiveness of auditors and information intermediaries, the determinants of manager's financial disclosure decision, the consequences of managers' financial disclosure decisions, the changing environment. Many researchers are interested in voluntary disclosures because this kind of reporting is unregulated. For example, Gu and Li (2007) examine stock price reaction to voluntary disclosure of innovation strategy by high-tech firms and its relation with insider stock transactions before the disclosure. Their evidence is consistent with insider purchase enhancing the credibility of the disclosure. And the pre-disclosure insider purchase is associated with greater future abnormal returns, suggesting that managers are privy to good news shortly before the disclosure.

With the ever-increasing use of the Internet, the Internet become a unique information disclosure tool that encourage flexible forms of presentation and allows immediate, broad, and inexpensive communication to investors. The majority of Internet disclosures is voluntarily and unregulated. Kelton and Yang (2008) examine the association between corporate governance mechanisms and disclosure transparency measured by the level of Internet financial reporting behavior. They developed a disclosure index to measure the extent of each sample firm's Internet financial reporting by presentation format, information content, and corporate governance disclosures.

Both of Gu and Li (2007) and Kelton and Yang (2008) extracting the information of disclosures are hand-collect and leave abundance information unexamined. And their results are likely to be influenced by the subjective judgment in the extracting information process.

Nowadays, we are living in an information age which is characterized by rapid growth in the amount of data collected and made available in electronic media. In order to process the flood of information, data mining is applied. Data mining is the process of identifying valid, novel, potentially useful, and ultimately understandable patterns in structure. The main difference between data mining and text mining is that in text mining, the patterns are extracted from natural language text rather than from structured databases of facts (Yang and Lee, 2005). There are about 85-90% of the corporate data stored in some sort of unstructured (text) form (McKnight, 2005). So many researches employ the text mining to explore the knowledge.

Some researches employ the text mining to investigate the customer relationship management. Chang, Lin and Wang (2008) apply the data warehouse and data mining technologies to analyze the customers' behavior in order to model owing to the enounced of the enounced of customer-orientation and making more effective marketing strategy. Lo (2008) employs the p-control chart to control the complaining rate by mining the customer complaints which include technical problems and non-satisfactory reports. Sakurai and Suyama (2005) employ text mining techniques to analyze e-mails collected at a customer center. Their results indicate that acquired concept relation dictionaries correspond to the intuition of operators in the customer center and give highly precise ratios in the classification.

Delen and Crossland (2008) using text mining to identify clusters and trends of related research topics form three major journals in the management information systems field. Basing on their findings,

this type of analysis could potentially be valuable for researchers in any field. Although the text mining is applied in many fields, it seems hard to find the paper applying text mining techniques in accounting fields.

There are many approaches to text mining, which can be classified form different perspectives, based on the inputs taken in the text mining system and the data mining tasks to be performed. In general, the major approaches are: (1) the keyword-based approach, where the input is a set of keywords or terms in the documents, (2) the tagging approach, where the input is a set of tags, and (3) the information-extraction approach, which inputs semantic information, such as events, facts, or entities uncovered by information extraction (Han and Kamber, 2006).

There are five steps for text mining:

a. File preprocessing: Creating a data set from the information collected.
b. Text parsing: Decomposing textual data and generating a quantitative representation suitable for data mining purposes.
c. Transformation (dimension reduction): Transforming the quantitative representation into a compact and informative format.
d. Document analysis: Performing clustering or classification of the document collection.
e. Results interpreting: Interpreting the results of text mining.

Data mining is a systematic approach to find underlying patterns, trends, and relationships buried in data and is regarded as a knowledge discovery method. Roiger and Geatz (2003) defined data mining as a process of employing one or more computer learning techniques to automatically analyze and extract knowledge out of data contained within a database. Data mining can be an automatic or semi-automatic process to discover and analyze volumes of data and find meaningful patterns or rules for many decision making problems (Berry and Linoff, 1997).

The researches regarding data mining can be classified into two categories: methodologies and techniques. The methodologies researches consist of data visualization, machine learning, statistical technique, and deductive database (Curt, 1995). The relevant applications of these methodologies include classification, prediction, clustering, summarization, linkage analysis, and sequential analysis (Fayyad et al., 1996). The techniques of data mining include statistical methods, neural networks, decision trees,

genetic algorithms, and non-parametric methods. The decision tree approach is another powerful classifier.

The classification analysis is a process for building a systematic classification model that establishes relationships between decision outcome and input variables. Several classification techniques have been proposed, including decision tree, neural network, nearest-neighbor classification, decision-rule induction, and Bayesian networks (Wei et al. 2003).

Neural Networks (NN) is a powerful tool to classify or cluster data. The theory of Neural Networks (NN) is originated in 1950s. At present, scholars brought several models of neural networks including Back-progagation Networks, Hopfield Networks and Radical basis Function Networks. An advantage of NN is not required to understand the framework of the model, like a black box, and people could utilize the tool easily. Some properties are not found in other data mining models, e.g. decision trees. These properties include robust performance in dealing with noisy or incomplete input pattern, high fault tolerance, and the ability to generalize from the input data (Patterson, 1996).

NN can be thought as a classifier. A classifier is a type of an inference engine that implements efficient strategies for computing subsumption relations between pairs of concepts, or for computing instance of relations between a concept and a set of instances. When it is used as a predictive model, it attempts to describe one column (the label) in terms of others (the attributes). A classifier is constructed from data where the label is known, and may be later applied to predict label values for new data where the label is unknown. For predicting the label values, a number of training examples which composed of a pair of a feature/variables and its label values (known) needs to be provided. The training set is used to train a classifier. After the classifier is trained, it is able to classify unlabeled values which we attempt to predict by the related feature/variables. Internally, a classifier is an algorithm or mathematical formula that predicts one discrete value for each input row and produce probability estimates for each of the label. NN relies on using input/output data patterns to train the network. After it has been adequately trained, it can provide appropriate responses to new come data (Hastie et al, 2001).

Another power tool is the decision tree approach which has been applied to many issues. Beynon et al. (2004) proposed a decision tree model to examine the relationship between firm characteristics and audit fees. Lee (2006) used the Classification And Regression Tree (CART) approach to explore the

performance of credit scoring. They presented four reasons of using CART as a research methodology. First, CART exhibits the capability of modeling complex relationships between variables without strong model assumption. Second, CART can identify “important” independent variables through the built tree when many variables are considered. Third, CART does not require much time for modeling and training process. Finally, the results of CART can be easily interpreted.

The CART has been applied to many research issues. Li (2006) applied CART to stainless steel production and found that the CART could produce insight to materials usage behaviors. Waheed et al. (2006) utilized CART to investigate the hyper-spectral remote sensing data to extract better crop information. Lee et al. (2006) employed CART to examine the customer credit of banks. Chang and Chen (2005) and Chang and Wang (2006) used CART to examine the risk factors associated with freeway traffic accidents and found that daily traffic volume is the most important determinant for freeway accidents.

Ⅲ、SEC news and Data

We collect the SEC news from the home page of U.S. Securities and Exchange Commission. There are 270 news released in 2007. We take the variation of S&P as the proxy of market reaction. Table 1 shows the descriptive statistics of S&P when SEC released a news. We find that there are 53% of observations their close index is higher than open index. Due to the limitation of Excel and data mining software, we cannot use all of the words in the news. Instead, we use the top 240 frequent words for data mining. Appendix 1 shows the top 240 frequent words and their occurrences.

Ⅳ Empirical results

We set a dummy variable as 1 if the close index of S&P is higher than open index of S&P. This study use CART as the classifier to find the rules. Table 2 demonstrates the decision rules. Figure 1 shows the variable importance. Table 3 shows the coincidence matrix. And table 4 shows the prediction rate. We can find that the prediction rate is 71.38%.

Table 1: the descriptive statistics of S&P

	open	high	low	close
average	1,478.44	1,487.05	1,468.49	1,477.71
maximun	1,564.98	1,565.42	1,555.46	1,565.15
minimun	1,377.86	1,388.09	1,363.98	1,374.12
std	45.06801	43.93611	44.71514	43.78662

Figure 1: The variable importance of CART experiment

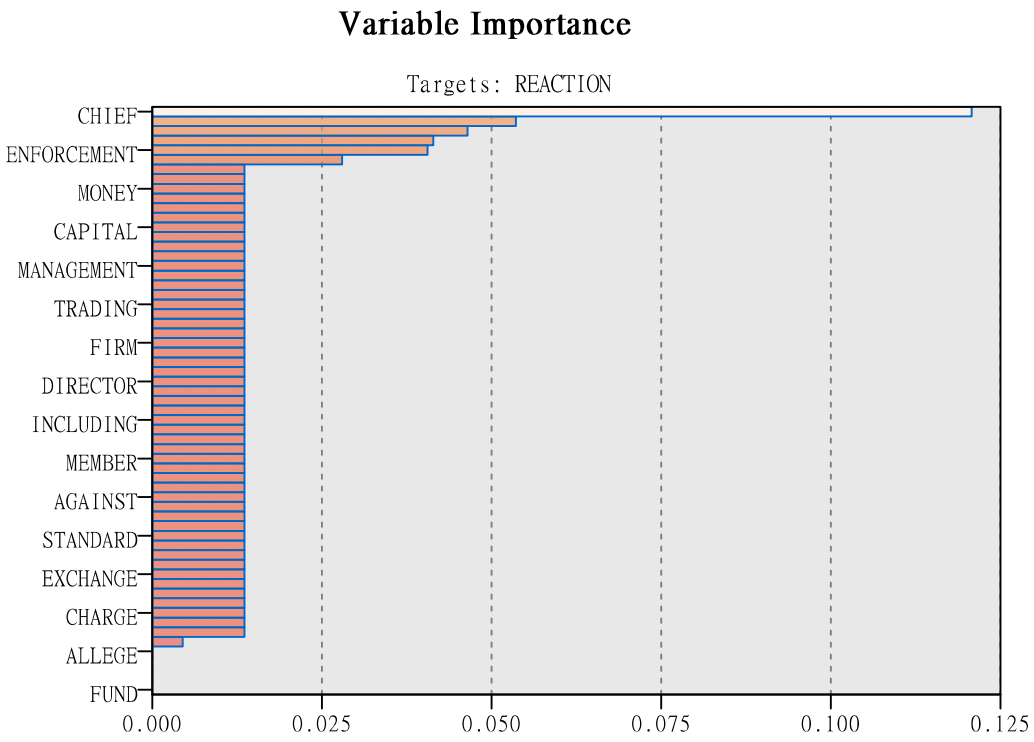

Table 2: decision rules
CHIEF <= 0.500 [Mode: 0]

```
        BROKER <= 1.500 [ Mode: 0 ]
            INTERACTIVE <= 1.500 [ Mode: 0 ] => 0
            INTERACTIVE > 1.500 [ Mode: 1 ]
                COMPANY <= 9.500 [ Mode: 1 ] => 1
                COMPANY > 9.500 [ Mode: 0 ] => 0
        BROKER > 1.500 [ Mode: 1 ]
            ACT <= 3.500 [ Mode: 1 ] => 1
            ACT > 3.500 [ Mode: 0 ] => 0
    CHIEF > 0.500 [ Mode: 1 ]
        SEC <= 6.500 [ Mode: 1 ]
            ACT <= 3.500 [ Mode: 1 ] => 1
            ACT > 3.500 [ Mode: 0 ] => 0
        SEC > 6.500 [ Mode: 0 ]
            COMMENT <= 0.500 [ Mode: 0 ]
                CAPITAL <= 0.500 [ Mode: 0 ] => 0
                CAPITAL > 0.500 [ Mode: 1 ] => 1
            COMMENT > 0.500 [ Mode: 1 ] => 1
```

Table 3: Coincidence Matrix (rows show actuals)

	0	1
0	111	14
1	63	81

Table 4 correct rate of decision tree

Correct	192	71.38%
wrong	77	28.62%

Reference

1. Berry, M.J., and G. Linoff,. 1997. Data mining techniques: for marketing sale and customer support, John Wiley and Sons, Inc., Canada.
2. Beynon, M. J. M. J. Peel and Yu-Cheng Tang. 2004. The application of fuzzy decision tree analysis in an exposition of the antecedents of audit fees. Omega 32:231-244.
3. Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. Classification and Regression Trees. Wadsworth.
4. Chang, Che-Wei, Chin-Tsai Lin and Lian-Qing Wang (2008). Mining the text information to optimizing the customer relationship management. Expert Systems with Applications, (in press)
5. Chang, L.Y. and H. W. Wang. 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accident Analysis and Prevention 38:1019-1027.
6. Curt, H. 1995. The devil's in the detail: techniques, tools, and applications for database mining and knowledge discovery-Part I. Intell. Software Strategies, 6:1-15.
7. Fayyad, U., G. Piatetsky-Sharpiro, P. Smyth. 1996. The KDD process for extracting useful knowledge from volumes of data. Communication ACM 39: 27-34.
8. Feldman, R. and Dagan, I. (1995). Knowledge discovery in textual database (kdt). In proceedings of the first international conference on knowledge discovery and data mining (KDD-95), Montreal, Canada, pp112-117.
9. Gu, Feng, and John Q. Li (2007). The credibility of voluntary disclosure and insider stock transactions. Journal of Accounting Research, 45(4), pp771-810.
10. Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). The elements of statistical learning : Data mining, inference, and prediction. New York: Springer.
11. Kelton, Andrea S., and Ya-we Yang. (2008). The impact of corporate governance on Internet financial reporting. Journal of Accounting and Public Policy, 27, pp62-87.
12. Kim, E., Kim W., and Lee, Y. 2003, Combination of multiple classifiers for the customers' purchase behavior prediction, Decision Support Systems 34(2): 167-175.
13. Lee, K., Booth, D., and Alam, P. (2005) A comparison of supervised and unsupervised neural

networks in predicting bankruptcy of Korean firms. Expert Systems with Applications 29:1-16.

14. Lee, T. S., Chiu, C. C., Lu, C. J., and Chen, I. F. (2002) Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications 23:245-254.
15. Lee, T.S., C.C. Chiu, Y.C. Chou and C.J. Lu. 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. Computational Statistics & Data Analysis 50: 1113-1130.
16. Patterson, D.W. (1996) Artificial Neural Networks: Theory and Applications. Englewood Cliffs, Prentice Hall.
17. Waheed, T., R.B. Bonnell, S.O. Prasher and E. Paulet. 2006. Measuring performance in precision agriculture: ACRT-A decision tree approach. Agricultural Water Management 84: 173-185.
18. Wei, C. S. Piramuthu, M. J. Shaw. 2003. Knowledge discovery and data mining, in C.W. Holsapple (Ed.), Handbook of Knowledge Management, vol. 2, Germany, Springer-Verlag, Berlin, Ch. 41: 157-18

Appendix 1: top 240 frequent words

word	occurrence	word	occurrence
commission	1628	registered	105
sec	1057	group	104
security	1035	associate	103
company	814	offering	103
investor	710	roundtable	103
financial	562	addition	102
exchange	530	make	102
fund	487	should	102
market	467	employee	100
office	428	linda	100
director	425	finance	99
information	420	related	99
million	404	effort	98
today	359	thomsen	98
said	349	include	97
public	345	registration	96
stock	343	system	96
option	338	web	96
complaint	336	according	94
enforcement	331	fraudulent	94
new	330	important	93
investment	316	before	92
act	301	date	92
against	299	made	92
washington	288	review	92
accounting	285	failed	91
division	277	member	91
law	271	period	91
year	271	revenue	91
order	269	risk	91
federal	262	university	91
rule	262	between	90
trading	253	corporate	90
allege	246	current	90
chairman	237	president	90
section	237	attorney	89
reporting	235	help	89

statement	214	actions	88
fraud	213	both	88
regional	213	credit	88
announced	201	interactive	88
under	201	firm	87
former	195	policy	87
scheme	194	required	87
state	193	three	87
compliance	189	benefit	86
district	181	proposed	86
transaction	180	part	85
chief	179	denying	84
fee	179	admitting	83
staff	179	amendments	83
including	176	assistance	83
penalty	174	served	83
disclosure	170	earnings	82
program	170	effective	82
management	169	form	82
cox	167	national	82
civil	166	provided	82
action	165	approximately	81
executive	165	industry	81
provision	162	orders	81
charge	157	paid	81
service	157	them	81
standard	156	advise	80
violation	156	auditor	80
pay	155	case	80
regulation	155	dollar	80
mutual	153	rate	80
sale	152	seminar	80
first	147	shareholder	80
business	146	settled	79
filed	146	accounts	78
www	145	activity	78
gov	144	antifraud	78
broker	143	issues	78
use	143	reports	78
share	142	using	78
provide	141	comment	77

without	140	general	77
practice	139	settlement	77
agency	136	chatman	76
bank	135	number	76
defendant	134	process	76
based	133	take	76
received	131	among	75
senior	130	asset	74
during	129	june	74
certain	127	payment	74
money	127	seniors	74
price	126	site	74
york	126	final	73
capital	124	march	73
day	124	programs	73
court	123	report	73
person	123	significant	73
time	122	additional	72
international	121	april	72
regulatory	121	conduct	72
christopher	120	consented	72
interest	120	officers	72
fair	119	regarding	72
disgorgement	118	return	72
officer	118	amount	71
grant	116	controls	71
corporation	114	cutomer	71
data	113	dealer	71
firms	112	development	71
further	112	income	71
available	111	issuerss	71
board	111	result	71
internal	111	subject	71
agreed	110	while	71
material	110	audit	70
compensation	109	other	70
http	109	own	70
investigation	109	private	70
record	109	profits	70
requirement	109	used	70
filing	108	plan	69

distribution	107	sept	69
united	107	well	69
each	106	committe	68
work	106	held	68
foreign	105	require	68

出席國際學術會議心得報告

計畫編號	NSC-98-2410-H-343-010
出國人員姓名 服務機關及職稱	洪嘉聲 南華大學會資系副教授
會議時間地點	99 年 7 月 29 日 至 99 年 8 月 3 日 美國/舊金山
會議名稱	2010 Annual Meeting of American Accounting Association
發表論文題目	Cash Holding Prediction: An Application of Two-layer Decision Trees

一、 參加會議經過

AAA 為全世界會計學界最重要的國際會議，所發表論文都相當嚴謹並具有創新性。會議中所發表的論文對實務界與學界均有相當的影響。2010 Annual Meeting of American Accounting Association 於美舊金山召開，此會議是每年舉辦一次，此次會議日期是從 7 月 29 日至 8 月 3 日止。此次所參與的會議場次是 8 月 1 日下午參加之口頭發表。

AAA 之論文均經過嚴格的評審審查，本人發表的論文名稱為: Cash Holding Prediction: An Application of Two-layer Decision Trees，報告過程很流暢，與美國、加拿大、韓國等國的教授一起討論。在討論中也獲得很多不一樣的想法，因此也讓我對於相關論文的撰寫方向與研究主題有許多啟示。

二、 與會心得

參加這次的國際會議，再到處觀摩他人海報及口頭報告的時候，往往無法一眼看透其他人研究的重點所在，只能盡量累積經驗，但發現此次發表論文的作者資質相當良莠不齊，有些人或許因為太緊張而導致在台上吞吞吐吐，有些人則能夠非常清楚明白的將自己的研究向聽眾分享。對我來說，以英文和外國學者交換心得是很新鮮的體驗，因為世界各地口音各各不同，有時候很難聽懂，有時候自己也會詞不達意，回國後必須要好好的加強英文，希望很快能夠再

度出國見世面長見識。此外，也利用會議空閒時間參觀舊金山市以及史丹福大學，增加許多見聞。

三、

參加本次會議一共攜回下列資料：

1. 大會論文集光碟片一張。
2. 大會論文簡介。

國科會補助計畫衍生研發成果推廣資料表

日期:2010/12/03

國科會補助計畫	計畫名稱: 資訊內容與市場反應：文字探勘之應用
	計畫主持人: 洪嘉聲
	計畫編號: 98-2410-H-343-010- 學門領域: 會計

無研發成果推廣資料

98 年度專題研究計畫研究成果彙整表

計畫主持人：洪嘉聲	計畫編號：98-2410-H-343-010-
計畫名稱：資訊內容與市場反應：文字探勘之應用	

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	1	30%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	0	1	15%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	30%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 □未發表之文稿 ■撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：(以 100 字為限)

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）(以 500 字為限)

本計畫利用資料探勘技術，探究 SEC 新聞對 S&P 指數的影響。此類研究尚屬少見。在研究過程中，對於資料之轉換與處理耗費極大的人力與時間，且受到 EXCEL 與資料探勘軟體之限制，所使用之變數僅能 250 個以下。對研究之限制頗多。