Natural Language Processing and Expert System Techniques for Equity Derivatives Trading: the IE-Expert System

Marco Costantino
Laboratory for Natural Language Engineering
Department of Computer Science
University of Durham, UK
Science Laboratories, South Road
Durham, DH1 3DL, UK
Tel. +44 181 932 2178, Fax. +44 181 932 2178
marco@afin.freeserve.co.uk

January 15, 1999

Abstract

Quantitative data are today largely analyzed by automatic computer programs based on traditional or artificial intelligent techniques, which provide traders with quantitative information that helps them hedge their risks. Qualitative data and, in particular, articles from on-line news agencies are instead not yet successfully processed. As a result, financial operators, notably traders, suffer from qualitative data-overload.

This paper describes how Natural Language Processing, Information Extraction and Expert Systems can be used for reducing the traders' qualitative information overload. In particular, the paper describes IE-Expert, an artificial intelligence system which is able to suggest investment decisions from qualitative information and to link this information to existing quantitative analysis.

Keywords: artificial intelligence, finance, NLP, information extraction, expert systems.

1 Introduction: Quantitative and Qualitative information

Equity derivatives traders have today access to a very large amount of information, both qualitative and quantitative, real-time and historical. Quantitative information consists of information which can be easily expressed in numbers (e.g. real-time prices from the major exchanges, volatility implied in exchange-traded option prices etc.). Qualitative information is instead information which cannot be easily expressed in numeric format, for example a sentence such as *"there are fears of an increase in the German interest rates"* from a news article. This includes prices of financial instruments quoted on any exchange e.g. equities, derivatives, exchange-rates, currencies etc.

The real-time quantitative information regarding the current behavior of the market together with the risk-management information (delta, gamma, vega, theta etc.) of the equity derivatives portfolios are used by traders to determine their trading strategies, aimed at maximizing their daily profits. In order to successfully determine their trading and hedging strategies, however, traders must have a *view* of the market. Let's assume for example that the traders' portfolios are positively correlated to movements of the underlyings in the portfolio. In this case, if the traders have a bearer view of the market and therefore expect the price of the underlying to drop, they will try to hedge their risks by making their portfolios inversely correlated to movements of the underlying price (negative delta). Traders make then use of the quantitative and risk information available to determine the quantity and quality of the hedge to put in place.

As we can see, quantitative information helps traders understand the risks associated to their position but, in the end, they must take a view of the market, based on their personal qualitative judgment. Traders must therefore take into account the qualitative aspects of the market and need to refer to the current qualitative information available which can be grouped into two main categories: **a)** analysis which is produced by internal analysts which includes forecasts for the main markets, indices and companies normally covering a period of one week; **b)** news from on-line news providers such as Dow Jones, Reuters, Bloomberg etc. which report the latest relevant news for the specific market or region. These news are the main source of qualitative information employed by traders to develop their view of the market. Figure 1 shows an example of the real-time news available to equity derivatives traders. A relevant number of articles are displayed every minute. The global-markets section of a generic trader's market sheet for example, tends to display an average of 5-7 articles a minute. Traders, often, are unable to capture and analyze this amount of information in such a short time and, therefore, the qualitative information is lost.

In this paper we focus on IE-Expert, a prototype system for processing financial qualitative information based on information extraction and expert systems technologies. In section 2 and 3 we introduce IE-Expert. Section 2 provides a general introduction to the system, while section 3 focuses on its implementation details. Finally, section 4 evaluates the results of the research.

2 IE-Expert: Integrating Information Extraction and Expert Systems

IE-Expert is based on information extraction and expert systems technology and is able to process these two categories of qualitative information and produce investment suggestions. In addition, it is able to provide a link between existing quantitative information and the investment decisions produced. The analysis is carried out in three main steps. The first step consists of the identification of relevant qualitative information from both real-time news and research material using the information extraction capabilities of IE-Expert. The second step consists of processing this information. Finally, the investment decision is shown to traders by linking it to existing quantitative information such as prices.

First of all, IE expert processes each of the incoming real-time news articles trying to identify any relevant information. The system processes all information using its information extraction

Quarto Group, the USM-traded publishing and printing services company, announced that it is buying Front Line Art Publishing, the California-based publisher of art prints and posters, for up to Dollars 9m (Pounds 6m). An initial payment of Dollars 7m will be satisfied by Dollars 5.3m cash and a Dollars 1.7m loan note. There is a further performance-related payment of up to Dollars 2m. For the 1993 year Front Line made profits of Dollars 1.4m, excluding owner remuneration, on turnover of Dollars 5m. Net assets at December 31 were Dollars 1.6m.

Template extracted by the system:

```
Template: TAKEOVER
    COMPANY_TARGET:   Front Line Art Publishing
    COMPANY_PREDATOR: Quarto Group
    TYPE_TAKEOVER:    FRIENDLY
    VALUE:            9 million dollars
    ATTRIBUTION:      Quarto Group
```

Figure 1: A template extracted from a financial news article.

capabilities. If no relevant information is found, the system skips the article and analyses the next incoming news. If any relevant information is found, a template is extracted according to the list of pre-defined templates shown in figure 2. Figure 1 shows a financial news article and the corresponding template extracted by the system. At this point IE-Expert retrieves from the database any information available regarding the two companies involved in the takeover and the market sector (according to the relevant region) which they belong to. For example, the system could retrieve the following information:

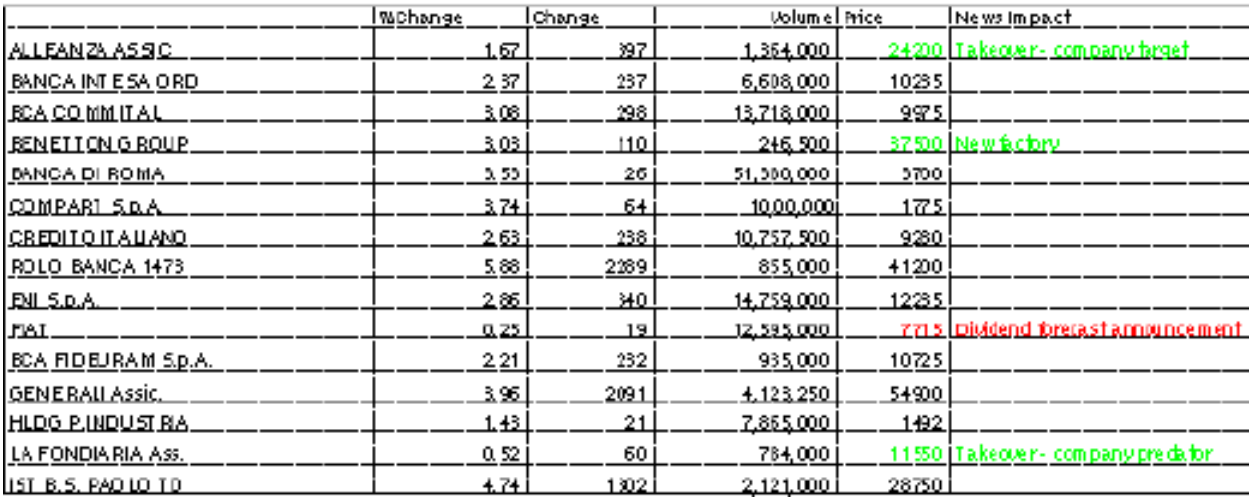
```
Company:       Tele-Communications Inc.
Negative:      Market under-performer

Company:       BELL ATLANTIC
Positive:      Buy

Market sector: American Telecommunications
Positive:      Expanding rapidly.
```

Once the relevant qualitative information has been identified and processed from both sources, it is fed to the financial expert system, which processes it according to specific rules and suggests an investment decision. The expert system's knowledge consists of a set of investment decision rules which match the financial templates shown in figure 2, which represent the most likely causes of share prices changes. From the information shown above, the expert system would produce the following investment suggestions:

```
1) BELL ATLANTIC:
                  Market Sector:   positive (expanding rapidly)
                  Company:         positive (buy)

                  Financial event: positive (takeover, company_predator)

                  Investment decision suggested: share price likely to
                  rise - buy

2) Tele-Communications Inc.
                  Market Sector:   positive (expanding rapidly)
                  Company:         negative (market under-performer)

                  Financial event: positive (takeover, company_target)
```

Company related	Company restructuring	General macroeconomics
Merger	New product	Interest rates movements
Takeover	Joint venture	Currency movements
Flotation	Staff changes	General macroeconomics data
New issue (shares, bonds etc.)	New factory	(inflation, unemployment
Privatization		trade deficit)
Market movement		
Bankruptcy		
Broker's recommendations		
Taking a stake		
Dividend announcement		
Overseas listing		
Profit/sales forecasts		
Profits/sales results		
Directors' dealings		
Legal action		
Investigation		

Figure 2: The pre-defined financial templates available in the system



	%Change	Change	Volume	Price	News Impact
ALLEANZA ASSIC	1.67	397	1,364,000	24200	Takeover- company target
BANCA INTESA ORD	2.37	237	6,608,000	10235	
BCA COMMITAL	3.08	298	13,718,000	9975	
BENETTON GROUP	3.03	110	246,500	37500	New factory
BANCA DI ROMA	3.53	26	51,300,000	3700	
COMPART S.p.A.	3.74	64	10,00,000	1775	
CREDITO ITALIANO	2.63	238	10,757,500	9280	
ROLO BANCA 1473	5.88	2389	855,000	41200	
ENI S.p.A.	2.86	340	14,759,000	12235	
FIAT	0.25	19	12,595,000	7715	Dividend forecast announcement
BCA FIDEURAM S.p.A.	2.21	232	935,000	10725	
GENERALI Assic.	3.96	2091	4,123,250	54900	
HLDG P.INDUSTRIA	1.43	21	7,865,000	1492	
LA FONDIARIA Ass.	0.52	60	784,000	11550	Takeover- company predator
IST B.S. PAOLO TO	4.74	1302	2,121,000	28750	

Figure 3: Merging real-time quantitative and qualitative information

```
Investment decision suggested: share price likely to
rise - buy
```

The expert system suggested a likely positive impact of the takeover event for both companies and that, as a consequence, the share price of the two companies is likely to increase.

The last step of the system is to display the investment decisions suggested by the expert system in real time to traders. This is done using a spreadsheet which reports the live quantitative information together with the information produced by the expert system. Figure 3 shows an example spreadsheet for part of the companies belonging to the MIB30 index. The color of the column "price" will change whenever a relevant news is processed by IE-Expert and an investment decision is suggested. The color will be red for events with negative impact on the share price, green, for events with positive impact and blue for events with no impact on the share price.

IE-Expert helps therefore traders overcome their qualitative data-overload and link the quantitive and qualitative information together, which allows them to quicker define their current view of the market for the next investment and hedging decisions.

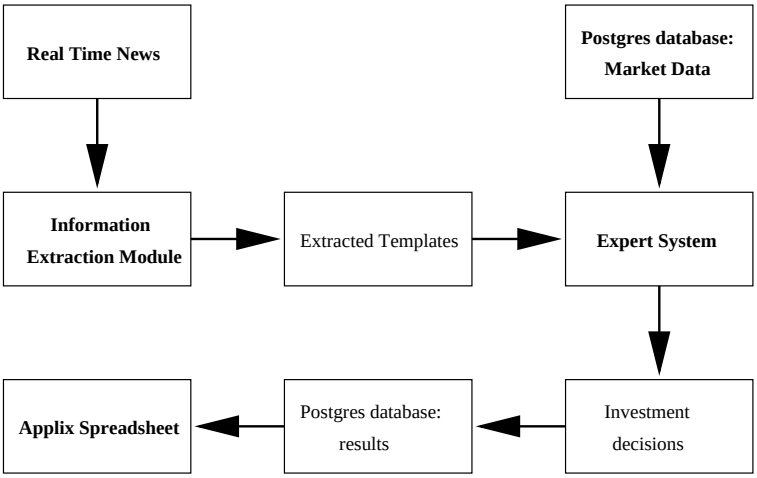


Figure 4: The architecture of IE-Expert

3 The implementation of IE-Expert

IE-Expert is based on two main components. The first component is the information extraction engine which identifies and extracts the relevant information from the incoming real-time financial news. The second component is the expert system, which is used to process the templates extracted by the information extraction component and the market analysis to produce investment suggestions. In addition, a postgres database stores the market data associated to a specific company or sector and an Applix spreadsheet is used to display the results. The architecture of the system is shown in figure 4.

The information extraction component is based on the Durham financial information extraction system, under development at the University of Durham, UK [Costantino, 1997]. The basic task of the natural language processing system is to process the input text and produce a representation of its meaning. This representation is then stored in an appropriate knowledge-base and can then be used for various different tasks and to generate natural language text. The core of the system is a large (over 100,000 nodes) semantic network, which consists of a hierarchy of nodes connected with arcs. The nodes represent entities (a company) and events (e.g. The company made a takeover). Each node is associated to specific control variables which are used to specify the type and properties of each node. Some of the control variables are as follows: **rank**: this control gives the nodes quantification; **type**: this control is very similar to grammatical qualifications **family**: this control groups the nodes into semantic "families" which share specific properties.

Source articles are processed by the system through four hierarchical modules: morphology, parsing, semantics and pragmatics (figure 5): The **morphology** module splits the input text into words and smaller units and produces for each word a list of possible meanings together with their syntactic and semantic categories. The input is then supplied to the parser; the **parser** module performs a full grammatical analysis of the source sentence recognizing the role of each of the words in the sentence, for example subject, object, verb and adjective. At this stage, the meaning of each of the words in the sentence is not yet determined, and will be resolved by the subsequent modules of the analysis; the **semantic analysis** module associates each of the words with their appropriate meanings and maps them onto the system's internal representation in a format compatible with the semantic network; the **pragmatic analysis** module performs the disambiguation of the meanings introduced by the semantic analysis module and type checking. At the end of the analysis process the new knowledge is stored in the semantic network. To produce the templates, the new knowledge obtained from the analysis of the source articles is matched against the templates definitions defined using the user-definable template interface, which is discussed in more detail in the next section and the final templates are extracted from the source texts.

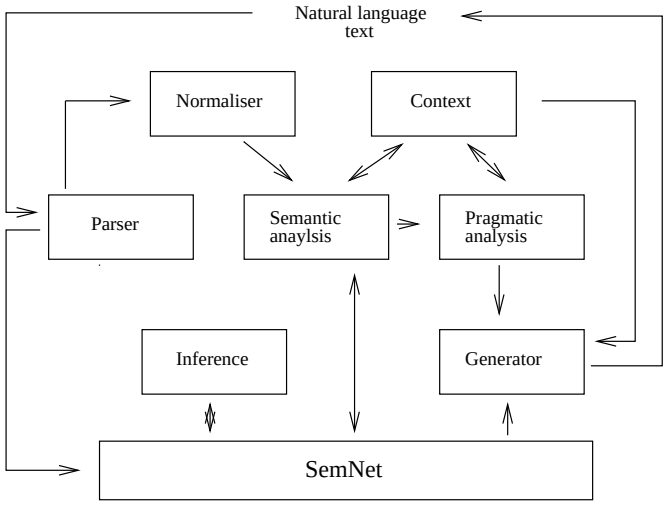


Figure 5: The Durham NLP System's core.

Once the templates have been produced, the system retrieves any associated market data information from the database which is currently based on a postgres server. The text of the template and the market data information are subsequently fed to the Expert System. The templates definitions and the market data available are parsed and stored in memory. At this stage, the information is processed using a set of rules corresponding to each of the pre-defined templates available in the information extraction module. The rule corresponding to the template is matched against the new information and an investment decision is produced. The expert system rules are represented as a table of True/False conditions which are matched against the slots of the template produced and of the associated market-data.

The investment decision produced by the expert system, together with the associated template and market information is subsequently stored in a database. Finally, an Applix Spreadsheet is used to retrieve the information from the database and display the results.

The information extraction system is written in the functional language Haskell and C language. The expert system is written in C. A postgres database is used for storing the market data information and the investment decisions and templates produced by the system. The information is displayed using an Applix spreadsheet which accesses directly the Postgres database. The system currently runs on a Sun SPARCstation with 80MB of RAM. However, it can easily be adapted for use within other Unix environments.

4 Evaluation and results

The evaluation of the results was carried out focusing on the performance of the information extraction module, which is essential for the system's success. This is because if any relevant information is missed or or non-relevant information is mistakenly extracted, the investment suggestions produced by the expert system could be misleading. The performance of the information extraction module was evaluated scoring the results of the information extracted for the user-defined takeover template from an evaluation set of 55 financial articles (25 relevant takeover articles and 30 non-relevant financial articles)[1]. Figure 1 shows a relevant takeover article from the evaluation set. The scores have been computed using the MUC-6 scoring program which was released to the developers of the MUC-6 systems [Chinchor and Dungca, 1995].

	P&R	2P&R	P&2R

[1] A complete discussion of the methodologies for evaluating information extractions systems is beyond the scope of this paper. More information regarding this topic can be found in [Chinchor and Dungca, 1995, Callaghan, 1998]

COMPANY_PREDATOR - investment suggestions		
Positive	**Neutral**	**Negative**
Company profile:positive Market sector: positive	Company profile: negative / negative Market Sector: positive or Company Profile: positive Market Sector: neutral / negative	Company profile: negative / neutral Market sector: negative / neutral

COMPANY_TARGET - investment suggestions		
Positive	**Neutral**	**Negative**
Company profile: positive / neutral / negative Market sector: positive positive / neutral / negative		

Figure 6: The expert systems rules for the takeover event.

```
F-MEASURES                    51.03     57.41     45.93

OVERALL PRECISION:            63%
OVERALL RECALL:               43%
```

The overall figure (51%) is rather high. The precision (63%) is significantly higher than the recall (43%). The high performance of the information extraction module should allow the expert system to produce correct investment suggestions.

5 Conclusions

In this paper we have shown how natural language processing, information extraction and expert systems can be used in finance. Information extraction and expert systems can be combined to process an incoming stream of news from on-line news providers, companies and market data to produce investment suggestions. The results can be subsequently linked to the existing real-time quantitative information. This allows traders overcome their qualitative-data overload and better define their trading and hedging strategies.

References

[Callaghan, 1998] P. C. Callaghan, *An Evaluation of LOLITA and Related Natural Language Processing Systems*, PhD thesis, University of Durham, Department of Computer Science, Laboratory for Natural Language Engineering, 1998.

[Chinchor and Dungca, 1995] N. Chinchor and G. Dungca, "The Scoring Method for MUC-6", in *Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann, November 1995.

[Costantino, 1997] M. Costantino, *Financial Information Extraction using pre-defined and user-definable templates in the LOLITA System*, PhD thesis, University of Durham, Department of Computer Science, Laboratory for Natural Language Engineering, November 1997.