Collect Meaningful Information about Stock Markets from the Web

Saleem Abuleil
Department of MMIS, Chicago State University
Chicago, IL, 60628, USA

and

Khalid Alsamara
Department of MMIS, Chicago State University
Chicago, IL, 60628, USA

ABSTRACT

Events represent a significant source of information on the web; they deliver information about events that occur around the world in all subjects and areas. These events can be collected and organized to provide valuable and useful information for decision makers, researchers, as well as for any person seeking knowledge. In this paper, we discuss an ongoing research to target stock markets domain to observe and record changes (events) when they happen, collect them, understand the meaning of each one of them, and organize the information along with meaning in a well-structured format. By using Semantic Role Labeling (SRL) technique, we have identified four factors for each event in this paper: verb of action and three roles associated with it, entity name, attribute, and attribute value. We have generated a set of rules and techniques to support our approach to analyze and understand the meaning of the events that take place in stock markets.

Keywords: Event Extraction and Understanding, Semantic Role Labeling, Stock Market, Arabic Language.

1. INTRODUCTION

Most businesses rely on the web to gather data that is crucial to their decision making processes. Data can be collected in many different ways, such as manual processes, which can ultimately prove to be inefficient and prone to errors, scripts and scrapers, which is a collection of scripts written by programmers, these "wrappers" query web sources as if they were databases, or automated web data monitoring and extraction which is the most cost-effective method for precise web extraction at scale.

Information Extraction (IE), as defined in the Message Understanding Conferences, has been traditionally defined as the extraction of information from a text in the form of text strings and processed text strings that are placed into slots labeled to indicate the kind of information that can fill them. The problem of extracting information from a large document collection can be approached using many different algorithms. The three classic models used in information extraction, (under which all these algorithms can be loosely grouped), are called Rule-based, Pattern Learning, and Supervised Learning.

Semantic parsing of sentences is believed to be an important task on the road to natural language understanding and has immediate applications in tasks, such as information extraction and question answering. Semantic Role Labeling (SRL) is a shallow semantic parsing task, in which for each predicate in a sentence, the goal is to identify all constituents that fill a semantic role, and to determine their roles (Agent, Patient, Instrument, etc.) and their adjuncts (Locative, Temporal, Manner etc.). For example, given a sentence like "the value of the index (Eiji X-30) increased to 9650 points", the task would be to recognize the verb "increased" as representing the predicate (verb), "(Eiji X-30)" as representing the entity, "price" as representing the attribute, and "$20.5" as representing the value of the entity in stock market. This is an important step towards making sense of the meaning of a sentence. In this paper we use Semantic Role Labeling (SRL) technique to extract information from text to be able to understand the meaning of it.

2. RELATED WORK

The last decade has seen great interest and many advances in the area of IE. In the US, the DARPA sponsored Tipster Text Program [14] and the Message Understanding Conferences (MUC) [20] have been the driving force for developing this technology. In fact, the MUC specifications for various IE tasks have become de facto standards in the IE research community. Therefore, it is necessary to present our IE effort in the context of the MUC program. MUC divides IE into distinct tasks, namely, NE (Named Entity), TE (Template Element), TR (Template Relation), CO (Co-reference), and ST (Scenario Templates) [12]. A variety of systems and techniques have been developed to address the information extraction problem. Successful techniques include statistical methods such as n-gram models, hidden Markov models, probabilistic context-free grammars [10], and rule-based methods that employ some form of machine learning. Rule-based methods have been especially popular in recent years. They use a variety of different approaches, but all recognize a number of common key facts.

Attia et al. [6] adapted and extended the automatic Multilingual, Interoperable Named Entity Lexicon approach to Arabic, using Arabic WordNet (AWN) and Arabic Wikipedia (AWK). First, they extract AWN's insatiable nouns and identify the corresponding categories and hyponym subcategories in AWK. Then, they exploit Wikipedia inter-lingual links to locate correspondences between articles in ten different languages in order to identify Named Entities (NEs). They apply a keyword search on AWK abstracts to provide for Arabic articles that do not have a correspondence in any of the other languages. In addition, they perform a post-processing step to fetch further NEs from AWK not reachable through AWN. Finally, they investigate discretization using matching with genomes databases, MADA-TOKAN tools and different heuristics for

restoring vowel marks of Arabic NEs. Using this methodology, they have extracted approximately 45,000 Arabic NEs.
Benajiba and Rosso [8] developed an annotated corpus (ANERcorp) collected from various news websites and the AWK. They also manually compiled gazetteers (ANERgazet) for location, people, and organization names that contained about 4,500 NEs. Shaalan and Raza [22] compiled gazetteers of NEs collected from annotated corpora, such as the ACE and ATB, from a database provided by government organizations and from Internet resources. The size of the database is presumably large, yet due to the extremely heterogeneous nature of the sources and the lack of a detailed taxonomy, it cannot be considered as a standard language resource. Similarly Benajiba et al. [9] tried to make up for the lack of Arabic NE lexical resources by including hand-crafted gazetteers for people, location, and organization names, and then semi-automatically enriched the location gazetteer using the AWK, taking the page labeled "Countries of the world" as a starting point to crawl AWK and retrieve location names. The resulting list went through manual validation to ensure quality.

Alkhalifa and Rodríguez [5] presented an approach to automatically attaching 3,854 Arabic NEs to English NEs using AWN, PWN, AWK and EWP as knowledge sources. Their approach is quite different as they start with an English NE collected from the PWN and EWP and try to obtain the Arabic counterpart from the AWK. Therefore they cannot capture Arabic NEs that have been originally compiled in Arabic and have no English equivalent. The AWK grows constantly and translation does not always keep pace.

Xiong et al. [24] integrated two discriminative feature-based models into a phrase-based Statistical Machine Translation (SMT) system, which used the semantic predicate-argument structure of the source language. Their first model defined features based on the context of a verbal predicate, to predict the target translation for that verb. Their second model predicted the reordering direction between a predicate and its arguments from the source to the target sentence.

Bazrafshan et al [7] proposed two methods for incorporating semantic role labels in a string-to-tree machine translation system by learning translation rules that are semantically enriched. In one approach, the system learned the translation rules by using a semantic role labeled corpus and augmenting the set of nonterminal used in the rules, and in the second approach, in addition to the regular SCFG rules, the system learned semantic roles which contained the complete semantic structure of a predicate, and added a feature to distinguish those rules.

Diab et al [13] presented an SRL system for Modern Standard Arabic that exploits many aspects of the rich morphological features of the language. The experiments on the pilot Arabic Propbank data showed that their system based on Support Vector Machines and Kernel Methods yields a global SRL F1 score of 82.17%, which improved the current state-of-the-art in Arabic SRL.

3. EVENTS EXTRACTION AND COMPREHENSION

In this paper, our target is stock markets domain on the web to observe and record changes (events) when they occur. We use Semantic Role Labeling (SRL) technique to find and extract those events, understand the meaning of each one of them, organize them according to their meaning and store them in a database. Each event has a verb and three roles associated with it: entity (proper name), attribute to describe the entity, and attribute value. Attributes could be presented in the sentence explicitly or implicitly.

We use the morphology systems [1] of nouns and verbs to analyze words in the text to generate their complete paradigms such as the number feature for nouns and verbs: singular, dual, and plural and tense feature for verbs: past and present progressives. We may find that more than one event appears in a single paragraph. To distinguish between them, we terminate the current event when we see either a new verb or a stop word. We have identified a set of cue words that signal a change of subject when they appear in the text.

Our approach looks for the events in the text, marks any paragraph that contains number(s), parses it, identifies the verb and the roles associated with it, extracts information about them, understands the meaning, and organizes them along with their meaning in a well-structured format. To understand the meaning of any event we identify its behavior and its attributes; in the next sections we discuss how we accomplish this task.

Parsing Rules
There are a number of combinations with respect to the way how the verb, entity, and attribute values (numbers) mentioned in the sentence. Table 1 shows all the scenarios and the chance of occurrence of each one of them.

TABLE 1: VERB, ENTITY AND VALUE COMBINATIONS

First Place	Second Place	Third Place	Frequency
Verb	Entity	Value	High
Verb	Value	Entity	Low
Entity	Verb	Value	Low
Entity	Value	Verb	Very Low
Value	Entity	Verb	
Value	Verb	Entity	

Based on combinations in table 1, we have generated the following rules to be used for parsing the text:

Event → Verb + W + Entity + W + SV + W + SE
| Verb + W + SV + W + Entity + W + SE
| Entity + W + Verb + W + SV + W+ SE
| Null

SE → Verb + W + SV + W + SE (co-reference)
| Entity + W + SV + W + SE (co-reference)
| Null

SV → V + W + SV | Null
V → N + {%} | N + {Point} | N + {Currency}
W → Extra zero, one or more words

*SE: Sub Event, SV: Sub Value, V: Value, N: Numeric Value.

Entities

Entities such as company name, bank name and stock market name are the core of events. Usually in Arabic language they come after keywords in the text such as index, share, and stock "مؤشر", "مزيج ", "سهم". We have collected number of keywords to use them to recognize and tag the entities in the events. We use some algorithms and techniques generated by [2, 3, 4] to find and tag the entities in the events.

Verbs (Event Behavior)

The verbs that we use to tag the events in the text represent the behavior of the events and they have the following features (past tense, singular, masculine/feminine) such as ارتفع "arose" انخفض "dropped" or a verb followed by either a positive noun or negative noun such as "اغلق … متراجعا" "closed … down" and "اغلق … مرتفعا" "closed … up". Based on the behavior of the verb in the sentence, we have classified it into two categories either positive when the verb indicates an increment in the value such as ارتفع "arose" or negative when the verb indicates a decrement in the value such as هبط "fall". A positive verb meaning the entity mentioned in the event is doing good and negative verb meaning the entity mentioned in the event is doing badly.

Numbers (Attribute Values)

Numbers mentioned in the events represent the values of the attribute that are used to describe those events. Numbers (attribute values) represent either a price (currency), worth (number of points) which is the value of a certain entity in stock market, or percentage of increment or decrement value (%). We have classified them according to two main factors: measuring unit and meaning.

Measuring units attached to numbers are three types: (1) currency symbol (sample: دولار / $ - Dollar), like " ارتفع 10 دولارات - Increased $10", (2) the word "نقطة" "Point" and its paradigms that represent the value of entity in stock market like "ارتفع ليصل 10 نقاط - Increased to reach 10 points" (3) and percentage (%) like "ارتفع بنسبة 10% - Increased by 10%".

Meaning of attribute values (numbers) can be identified by some particles located between the numbers and the verbs in the events. Particles such as {الى "to", ل "to" ب "by", عند "at", على "on"} determine the meaning of the attribute values (numbers) in the events; we have generated two rules to cover this aspect as follows:

Rule #1: govern price value or worth value

$$\text{Verb} + W_{\{0..n\}} + \{\text{على | عند | ل | الى}\} + W_{\{0..n\}} + \text{Value}$$

Examples:

Increased to reach 10 points	ارتفع ليصل 10 نقاط
Increased to 10 points	ارتفع الى 10 نقاط
Increased and reached at $10	ارتفع ووصل عند 10 دولار
Increased and closed on $10	ارتفع ليغلق على 10 دولار

Rule #2: govern Increment/Decrement Value in Price or Worth Increment/Decrement Percentage in Price or Worth

$$\text{Verb} + W_{\{0..n\}} + \{\text{ب} \mid \text{Null}\} + W_{\{0..n\}} + \text{Value}$$

Examples:

Increased $10	ارتفغ 10 دولارات
Increased by 10%	ارتفع بنسبة 10%
Increased by the of value $10	ارتفع بمقدار 10 دولارات

Attributes

Attributes represent the main source of information about entities in the events. We have identified eight different types of attributes to help us describe the events and understand the meaning of each one of them. We use three factors to determine which attribute to use in each case: the measuring unit of the number (currency, points, and percentage), the meaning of the number, as demonstrated in the previous section and the verb type (behavior). For example to use the attribute "price-increment", we have to satisfy three conditions: number meaning (rule #2) to indicate increment/decrement in value/percentage, number measuring unit is currency, and verb type (behavior) is positive. Attributes could be presented in the sentence explicitly or implicitly. See table 2.

TABLE 2: ATTRIBUTES

Attribute	Number Meaning [Rule #]	Number Unit	Verb Type (behavior)
Price	1	Currency Symbol	N/A
Worth	1	Point(s)	N/A
Price-Increment	2	Currency Symbol	P
Price Decrement	2	Currency Symbol	N
Worth-Increment	2	The word "Point" and its paradigms	P
Worth-Decrement	2	The word "Point" and its paradigms	N
Percentage – Increment	2	%	P
Percentage - Decrement	2	%	N

P: Positive, **N**: Negative, N/A: Not Applicable

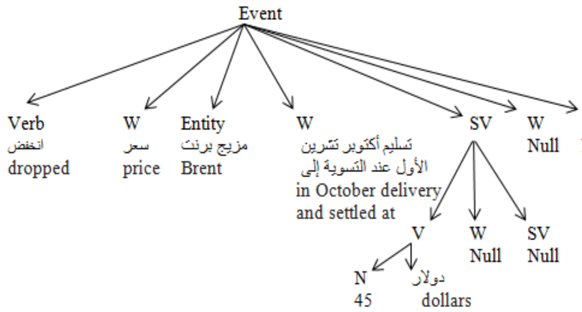
Examples

Here are some examples to illustrate how the Events Extraction and Understanding approach works and how it identifies the meaning of the events and organize the information about them. Our approach is able to understand the meaning of the extracted events and to know if a certain entity in a specific event is doing good or not, it is able to tell if a certain entity is losing or gaining in its value and by what percentage, it understands the meaning of the concepts losing points or gaining points, losing in price or gaining in price, and increment/decrement in price and value of a certain entity in stock market.

Example 1

وانخفض سعر مزيج برنت تسليم أكتوبر تشرين الأول عند التسوية إلى 45 دولار للبرميل

Brent price dropped in October delivery and settled at 45 dollars a barrel

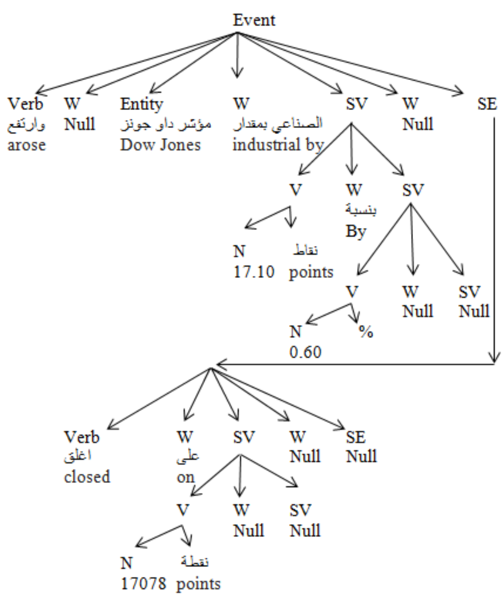


<Event>
<Verb: dropped انخفض>
<Behavior> negative </Behavior>
</Verb>
<Entity> Brent مزيج برنت</Entity>
<Attribute: Price >
<Value> 45 dollars دولار 45</Value>
</Attribute>
</Event>

Example 2

وارتفع مؤشر داو جونز الصناعي بمقدار 72.10 نقاط بنسبة 0.06 % واغلق على 17078 نقطة

The industrial index Dow Jones arose 72.10 points, by 06.0 % to close at 17078 points

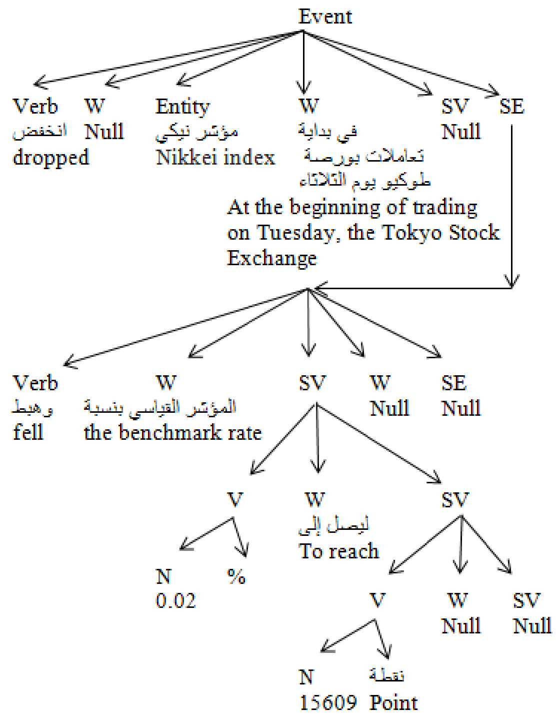


<Event>
<Verb: arose ارتفع>
<Behavior> positive </Behavior>
</Verb>
<Entity> Dow Jones مؤشر داو جونز </ Entity>
<Attribute-1: Worth-Increment>
<Value> 17.10 point نقطة 17.10</Value>
</Attribute-1>
<Attribute-2: Percentage-Increment>
<Value> 06.0 %</Value>
</Attribute-2>
<Attribute-3: Worth>
<Value> 17078 points نقطة 17078</Value>
</Attribute-3>
</Event>

Example 3

انخفض مؤشر نيكي القياسي في بداية تعاملات بورصة طوكيو يوم الثلاثاء, وهبط المؤشر القياسي بنسبة 0.02 % ليصل إلى 15609 نقطة

The Nikkei index dropped in early trading on Tuesday, the Tokyo Stock Exchange, the benchmark index fell by 0.02% to 15609 points



<Event>
<Verb dropped انخفض - fell وهبط>
<Behavior>negative</Behavior>
</Verb>
<Entity> Nikkei index مؤشر نيكي</Entity>
<Attribute-1: Percentage-Decrement>
<Value> 2% </Value>
</Attribute-1>
<Attribute-2: Worth>
<Value> 15609 points نقطة 15609</Value>
</Attribute-2>
</Event >

4. Stock Market Trends

The information we extract about stock markets can be used to support different types of applications such as question-answering systems, summarization systems, and in studying stock markets trends. The focus in this paper is on stock market trends to present them in two different ways: the behavior of a certain stock or index, show changes in values and prices of a certain stock or index and make a comparison between different stocks.

Example:
In the following example we demonstrate how the system uses the extracted information to organize it to generate trends. In table 3 and table 4 we show the data we extracted for two stocks TecDax and Stoxx50 for a certain date for a certain period of time, we record the events every one hour, we sort it in ascending order by time, the

last column represents the final value of the stock at the end of each event. In figure 1 the system generates a trend to show the recorded values in the extracted events for Stoxx50 and in figure 2 the system generates a trend to show the recorded values in the extracted events for TecDax. Figure 3 and figure 4 show the behavior of the stocks, we use the value 1 to represent positive behavior and value 0 to represent negative behavior, and figure 5 shows a comparison in behavior between both stocks.

Table 3: TecDAX

time	Behavior	Attribute	Attribute Value	Final Value (computed)
9:10	Opening	---------	---------	1500.01
10:00	positive	Value-Increment	3.93	1503.94
11:00	Negative	Value -Decrement	5.36	1498.58
12:00	Negative	Value	1494.50	1494.50
13:00	Positive	Value-Increment	1.33	1495.83
14:00	Negative	Percentage - Decrement	0.167%	1493.33
15:00	Negative	Value -Decrement	0.096%	1491.89
16:00	Positive	Percentage - Increment	0.422%	1498.20
17:00	Negative	Value -Decrement	1.8	1496.40

Table 4: Stoxx 50

time	Behavior	Attribute	Attribute Value	Final Value (computed)
9:10	Opening	----------	---------	3,389.97
10:00	Negative	Value - Decrement	11.73	3,378.24
11:00	Negative	Value -Decrement	9.09	3,369.15
12:00	Negative	Percentage - Decrement	0.469%	3,353.32
13:00	Positive	Value-Increment	17.49	3,370.81
14:00	Negative	Percentage - Decrement	0.342%	3,359.28
15:00	Negative	Value	3,344.86	3,344.86
16:00	Positive	Percentage - Increment	0.516%	3,362.13
17:00	Negative	Percentage - Decrement	0.338%	3,350.74

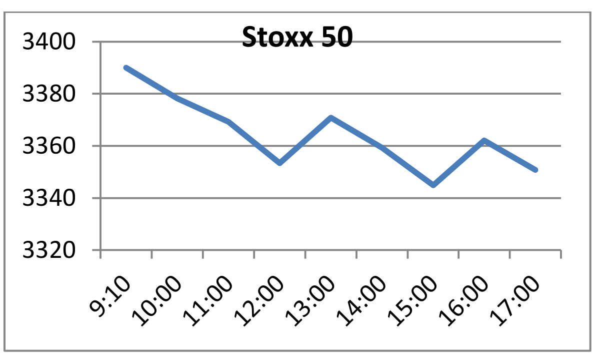


Figure 1: Stoxx50 Value Trend

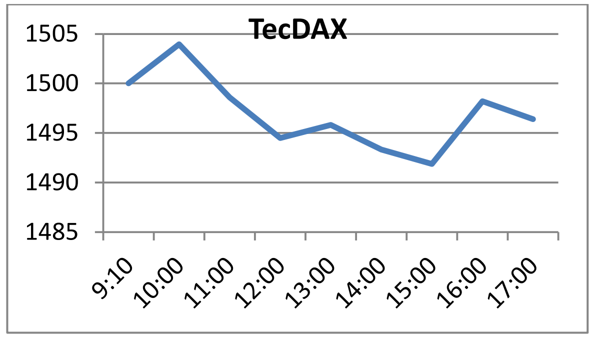


Figure 2: TecDAX Value Trend

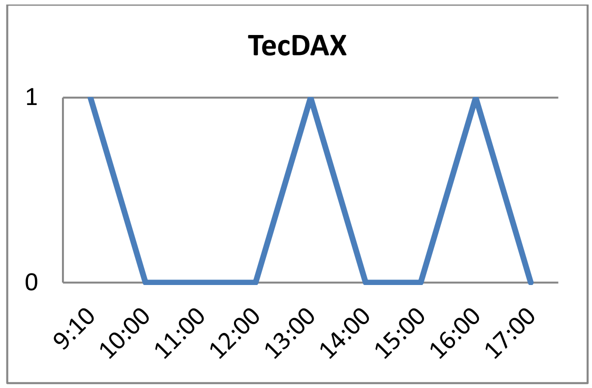


Figure 3: TecDax Behavior Trend

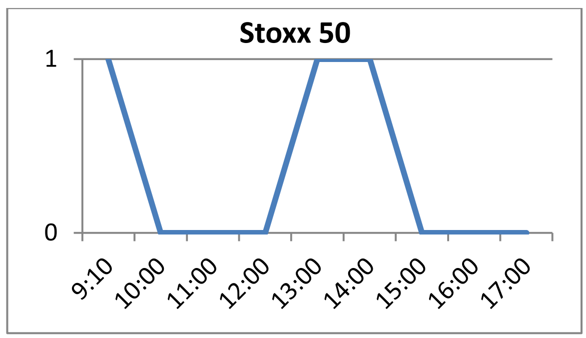


Figure 4: Stoxx50 Behavior Trend

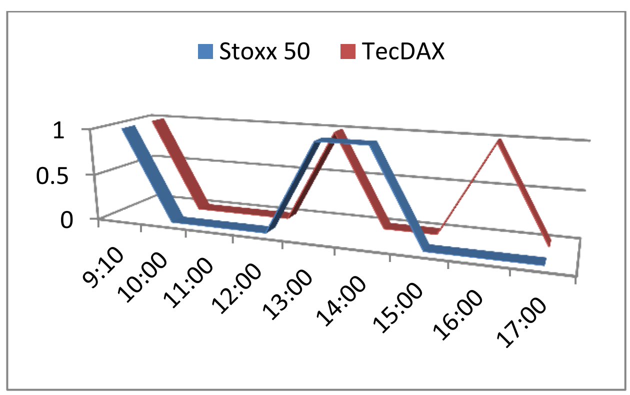


Figure 5: Comparison between Stoxx50 and TecDax

5. Experiments and Analysis

We have tested our approach manually on 200 events, from four different websites specialized in stock exchange and market news [16, 17, 18, 19]. First we collected 110 different entity names mentioned in the events, second we identified and collected the verbs and nouns that we use to tag the events in the text and third we classified each verb/noun to either positive or negative. A positive verb, meaning the entity mentioned in the event, is doing good and the negative verb, meaning the entity mentioned in the event, is doing badly. See table 5.

TABLE 5: SAMPLE OF VERBS

Verb (Arabic)	Verb (English)	Frequency	Type (Behavior)
تراجع\ت	Retreated	71	Negative
ارتفع\ت	Arose	51	Positive
صعد\ت	Stepped up	21	Positive
انخفض\ت	Dropped	16	Negative
هبط\ت	Declined	16	Negative
زاد\ت	Increased	12	Positive
خسر\ت	Lost	8	Negative
اغلق\ت + Positive Noun	Closed + Positive Noun	8	Positive
ربح\ت	Won	7	Positive
اغلق\ت + Negative Noun	Closed + Negative Noun	6	Negative

After collecting and preparing the data we tested our approach against the events to collect the information about each event and organize it with respect to the meaning of each one of them. As a whole, the results were as follows: out of 200 events, our approach parsed 155 events correctly, 27 events incorrectly, and missed 18 events. The Recall is 89.5%, the Precision is 85% and Slot Error Rate (SER) is 26%.

We use two rules to classify attribute values (numbers) according to their meaning in the events based on the particles associated with each one of them. The precision of the first rule is 96.5% and the precision of the second rule is 94%. For more details see table 6. Attributes represent the main source of information about entities in the events. We have identified eight different attributes to help us understand the meaning of the events. Our approach identified 93% of the attributes correctly. We have collected a number of garbage words to ignore them in the future to improve the accuracy of our approach. Table 7 shows some of them with the frequency of each one of them.

The source of both missing events and the event that are parsed incorrectly can be summarized as following:

[1] Increment/decrement value is not mentioned explicitly in the sentence. Example:

وخسر مؤشر توبكس الأوسع نطاقا 0.3 % مسجلا 1293.21 نقطة بينما انخفض مؤشر " جيه . بي . اكس – نيكي 400 " **<u>بنفس النسبة</u>** إلى 11732.03 نقطة
The Topix index lost 0.3% range, recording 1293.21 points, while the G.B. X. - Nikkei 400 fell in **<u>the same proportion</u>** to 11732.03 points

[2] Event spreads into two long sentences and it loses connection between its factors.

[3] Verb is missing or not previously identified. Example:

<u>تصدرت</u> فلسطين لتمويل الرهن العقارى الشركات المرتفعة بنسبة 4.82%

<u>Topped</u> Palestine Mortgage and Real Estate High Companies by 4.82 %

[4] Verb is in present tense format and not past tense format. Example:

وتصدر قطاع البنوك الأوروبي القطاعات الخاسرة بعد **<u>انخفاض</u>** مؤشر ستوكس-600 2.8 %
The European banks sector leads the losing sectors after the **<u>reduction</u>** of index Stoxx-600 2.8%

[5] Compare values from two different periods in the same sentence. Example:
وصعد مؤشر "يوروفرست 300" لأسهم الشركات الأوروبية الكبرى 1.1 % عند الإغلاق إلى 1399.43 نقطة **<u>بعدما قفز لأعلى مستوى له منذ مطلع العام 2008م عند 1400.99 نقطة</u>**.

The Ftseurofirst-300 index ascended of top European shares 1.1% at the close to 1399.43 points, **<u>having jumped to its highest level since the beginning of the year 2008 at 1400.99 points</u>**

TABLE 6: NUMBER MEANING RULES

Rule #	Particle	# Correct	# Incorrect
1	ب "by"	187	0
1	null	63	9
2	الى "to"	73	3
2	عند "at"	28	1
2	ل "to"	5	2
2	على "on"	5	1

TABLE 7: SAMPLE OF GARBAGE WORDS

Word (Arabic)	Word (English)	Frequency
سجل	Recorded	54
بلغ	Reached	35
وصل	Reached	21
قياسي	Standard	15
سوق	Market	13
استحوذ	Hauled	10
حقق	Achieve	9
عام	General	8
تباين	Contrast	7

6. CONCLUSION

In this paper we discussed an ongoing research in the Arabic language to understand the meaning of stock market events using Semantic Role Labeling (SRL) technique. We have generated a set of rules and techniques to parse the events, analyze them, recognize their behavior and identify their attributes to understand the meaning of each one of them. Although our approach is targeting the Arabic language, we believe it can be applied to other languages as well.

7. REFERENCES

[1] Abuleil, S. and Evens, M., 2002. Extracting an Arabic Lexicon from Arabic Newspaper Text. **Computers and the Humanities**, 36(2), pp. 191-221.

[2] Abuleil, Saleem, 2004. "Extracting Names From Arabic Text for Question-Answering Systems". RIAO'04, **Proceeding of the 7th International Conference on Coupling Approaches, Coupling Media, and Coupling Languages For Information Retrieval. University of**

Avignon (Vaucluse), France April 26th-28th, 2004. pp 638-647.
[3] Abuleil, Saleem and Evens, Martha, 2004. “Named Entity Recognition and Classification for Text in Arabic”. IASSE’04, **Proceedings of the 13th International conference in Intelligent Systems and Software Engineering**. Nice, France, July 1-3, 2004. pp 89-94.
[4] Abuleil, Saleem and Khalid Alsamara 2006, “Refining Topics in Named-Entity Recognition and Classification for Text in Arabic”, **Proceeding of the 6th International Conference in Information Technology (ICIT’06),** Irbid, Jordan, December 19-21, 2006.
[5] Alkhalifa, M., and Rodr´ıguez, H., 2009. Automatically extending ne coverage of arabic wordnet using wikipedia. In **CITALA**’09, Rabat, Morocco.
[6] Attia, M., Toral, A., Tounsi, L., Monachini, M., and Genabith J. V., 2010. An Automatically Built Named Entity Lexicon for Arabic. Valletta, Malta, May 19-21, 2010. **LREC** 2010 proceedings.
[7] Bazrafshan, M. and Gildea, D, 2013, Semantic Roles for String to Tree Machine Translation, **The 51st Annual Meeting of the Association for Computational Linguistics** Short Papers (ACL Short Papers 2013**)**, Sofia, Bulgaria, August 4-9, 2013
[8] Benajiba, Y., and Rosso, P., 2007. Anersys 2.0 : Conquering the ner task for the Arabic Language by combining the maximum entropy with pos-tag information. In **IICAI-**2007, Pune, India.
[9] Benajiba, Y., Diab, M., and Rosso, P., 2008. Arabic named entity recognition using optimized feature sets. In **EMNLP-**2008, Honolulu, Hawaii.
[10] Califf, M. E., & Mooney, R. J. (1999). Relational Learning of Pattern-Match Rules for Information Extraction. In Proceedings of **the Sixteenth National Conference on Artificial Intelligence (AAAI99)**, Orlando, FL, pp. 328-334.
[11] Charles L. Wayne Multilingual topic detection and tracking: Successful research enabled by corpora and evaluation., 2000, . In LREC 2000: **2nd International Conference on Language Resources and Evaluation**, Athens, Greece, June 2000.
[12] Chinchor, N., and Marsh, E. (1998) MUC- 7 Information Extraction Task Definition (version 5.1), In Proceedings of **MUC-7**.
[13] Diab, Mona, Alessandro Moschitti, and Daniele Pighin. "Semantic Role Labeling Systems for Arabic using Kernel Methods."**Proceedings of ACL-08: HLT,** June (2008): 798-806
[14] Grishman R. (1997) **TIPSTER Architecture Design Document Version 2.3. Technical report, DARPA**.
[15] Hagen F¨urstenau and Mirella Lapata. 2012. Semi-supervised semantic role labeling via structural alignment. **Computational Linguistics**, 38(1):135– 171.
[16] http://www.fxnewstoday.ae/arab-market.
[17] http://www.mubasher.info.
[18] http://www.souqelmal.com.
[19] http://www.argaam.com.
[20] Khandelwal, V., Gupta, R. and Allan., J. (2001). An evaluation corpus for temporal summarization. In James Allan, editor, Proceedings of HLT 2001, **First International Conference on Human Language Technology Research**, San Francisco, 2001, Morgan Kaufmann.
[21] MUC-7 (1998) **Proceedings of the Seventh Message Understanding Conference (MUC-7),** published on the website _http://www.muc.saic.com.
[22] Shaalan, K., and Raza, H., 2009. Nera: Named entity recognition for Arabic. In **JASIST**, John Wiley and Sons, pages 1652–1663, NJ, USA.
[23] Tagyoung Chung, Licheng Fang, and Daniel Gildea. 2011. Issues concerning decoding with synchronous context-free grammar. **In Proceedings of the ACL 2011 Conference** Short Papers, Portland, Oregon. Association for Computational Linguistics.
[24] Xiong, D., Zhang, M., and Li, H. 2012. Mod- eling the translation of predicate-argument structure for smt. In **ACL** (1), pages 902–911.