An Overview of Event Extraction from Text

Frederik Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska de Jong

Erasmus University Rotterdam
PO Box 1738, NL-3000 DR
Rotterdam, the Netherlands

{fhogenboom, frasincar, kaymak, fdejong}@ese.eur.nl

Abstract. One common application of text mining is event extraction, which encompasses deducing specific knowledge concerning incidents referred to in texts. Event extraction can be applied to various types of written text, e.g., (online) news messages, blogs, and manuscripts. This literature survey reviews text mining techniques that are employed for various event extraction purposes. It provides general guidelines on how to choose a particular event extraction technique depending on the user, the available content, and the scenario of use.

1 Introduction

With the increasing amount of data and the exploding number of digital data sources, utilizing extracted information in decision making processes becomes increasingly urgent and difficult. An omnipresent problem is the fact that most data is initially unstructured, i.e., the data format loosely implies its meaning [9] and is described using natural, human-understandable language, which makes the data limited in the degree in which it is machine-interpretable. This problem thwarts the automation of for example vital information retrieval (IR) and information extraction (IE) processes – used for decision making – when involving large amounts of data.

Text Mining (TM) [15] is concerned with information learning from preprocessed text (e.g., containing identified parts of speech or stemmed words). By means of text mining, often using Natural Language Processing (NLP) [22] techniques, information is extracted from texts of various sources, such as news messages and blogs, and is represented and stored in a structured way, e.g., in databases. A specific type of knowledge that can be extracted from text by means of TM is an event, which can be represented as a complex combination of relations linked to a set of empirical observations from texts.

An example of an event is an acquisition. If we consider the representation `<Company> <Buy> <Company>`, words identified in text referring to companies are linked to the concept `<Company>`, and (conjugations of) verbs having the meaning of acquisition are associated with `<Buy>`. Representations of this event can be extracted from news message headers such as "*Google acquires Picnik*", "*Lala bought by Apple*", or "*Skype sold to Microsoft*".

Event extraction from unstructured data such as news messages could be beneficial for IE systems in various ways. For instance, being able to determine events could enhance the performance of personalized news systems [2, 10], as news messages can be selected more accurately, based on user preferences and identified topics (or events). Furthermore, events can be useful in risk analysis applications [3], monitoring systems [17], and decision making support tools [36].

Extracted events are also extensively applied within the medical domain [6, 38], where event parsers are utilized for extracting medical or biological events like molecular events from corpora. Another possible – but less researched – application of event extraction lies within the field of algorithmic trading, representing the use of computer programs for entering trade orders with algorithms deciding aspects like timing, price, and quantity of an order. Financial markets are extremely sensitive to breaking news [24]. Economic events like mergers and acquisitions [31], stock splits [14], dividend announcements [23], etc., play a crucial role in the daily decisions taken by brokers, where brokers can be humans or machines. Besides being able to process news faster, machines are able to deal with larger volumes of emerging news, having access to more information than we humans do, and thus making better informed decisions.

Given the promising potential for applications of event extraction, and assuming that the challenges of real-time extraction and combination of events can be tackled adequately, it is worthwhile to investigate which text mining techniques are appropriate for this purpose. The current body of literature is lacking a high-level survey on event detection in text. Therefore, the goal of this paper is to review existing approaches to event extraction from text. We aim for providing general guidelines on selecting the proper text mining techniques for specific event extraction tasks, taking into account the user and its context. For this, we strive for a similar overview of performance aspects and recommendations as has been developed for cross-lingual research systems [25]. The work presented herein is a first step, focussing specifically on event extraction from text. The recognition of the space and time event dimension in text is considered outside the scope of this paper.

Throughout this paper we evaluate event extraction approaches using several criteria. For this, we review data that are available in the literature and distinguish between the categories high, medium, and low. First of all, we investigate the amount of data needed for each approach. Moreover, the amount of required domain knowledge is evaluated, together with the required amount of user expertise. Finally, we also discuss the interpretability of the results.

This paper continues with an elaboration of approaches to event extraction in Section 2. Subsequently, Section 3 presents a discussion on the event extraction approaches introduced in this survey. Finally, Section 4 concludes the paper.

2 Event Extraction

We distinguish between three main approaches to event extraction, in analogy with the classic distinction that is made in the field of modeling. First, there are data-driven approaches, described in Section 2.1, which aim to convert data to knowledge through the usage of statistics, machine learning, linear algebra, etc. Second, we distinguish expert knowledge-driven methods as discussed in Section 2.2, which extract knowledge through representation and exploitation of expert knowledge, usually by means of pattern-based approaches. Finally, the hybrid event extraction approaches elaborated on in Section 2.3 combine knowledge and data-driven methods.

2.1 Data-Driven Event Extraction

Data-driven approaches are commonly used for natural language processing applications. These approaches rely solely on quantitative methods to discover relations. Data-driven approaches require large text corpora in order to develop models that approximate linguistic phenomena. Furthermore, data-driven text mining is not restricted to basic statistical reasoning based on probability theory, but encompasses all quantitative approaches to automated language processing, such as probabilistic modeling, information theory, and linear algebra.

One could distinguish between many approaches, e.g., word frequency counting, ranking by means of the Term Frequency – Inverse Document Frequency metric, word sense disambiguation, n-grams, and clustering. Despite their differences, all approaches focus on discovering statistical relations, i.e., facts that are supported by statistical evidence. Examples of discovered facts are words or concepts that are (statistically) associated with one another. However, statistical relations do not necessarily imply semantically valid relations, nor relations that have proper semantic meaning.

Several examples of the usage of data-driven text mining approaches for event extraction can be found in literature. For instance, in their 2009 work, Okamoto et al. [27] elaborate on a framework for detection of occasional or local events, which employs hierarchical clustering techniques. While clustering itself could already yield promising results for event extraction, the authors of [21] make use of a combination of weighted undirected bipartite graphs and clustering in order to extract key entities and significant events from daily web news. Clustering techniques are also employed by Tanev et al. [34], who also aim for real-time news event extraction, but focus especially on violence and disaster events. The authors make use of automatic tagging of words and the presented framework is designed to automatically learn patterns from discovered events. Lastly, the authors of [19] also employ word-based statistical text mining in their work from 2005. The authors elaborate on a framework aimed at news event detection, based on support vector machines.

A drawback of the discussed data-driven methods to event extraction is that they do not deal with meaning explicitly, i.e., they discover relations in corpora without considering semantics. Another disadvantage of statistics-based text mining is that a large amount of data is required in order to get statistically significant results. However, since these approaches are not based on knowledge, neither linguistic resources, nor expert (domain) knowledge are required.

2.2 Knowledge-Driven Event Extraction

In contrast to data-driven methods, knowledge-driven text mining is often based on patterns that express rules representing expert knowledge. It is inherently based on linguistic and lexicographic knowledge, as well as existing human knowledge regarding the contents of the text that is to be processed. This alleviates problems with statistical methods regarding meaning of text. Information is mined from corpora by using predefined or discovered linguistic patterns, which can be either lexico-syntactic patterns [11, 12] or lexico-semantic patterns [2]. The former patterns combine lexical representations and syntactical information with regular expressions, whereas the latter patterns also make use of semantic information. Semantics are usually added by means of gazetteers, which use the linguistic meaning of text [7, 8], or by means of ontologies [10, 32].

Several attempts have been made for extracting events using pattern-based approaches to text mining. Both – mostly manually created – lexico-syntactic and lexico-semantic patterns are used; the former more often than the latter. For instance, in their 2009 work, Nishihara et al. [26] extract personal experiences from blogs by means of three keywords (place, object, and action) that together describe an event. For this, sentences are split using lexico-syntactic patterns. A similar approach can be found in [1], where the authors focus on pattern-based relation and event extraction. Here, lexico-syntactic patterns are employed in order to discover a wide range of relations and events in the domains of finance and politics. The authors of [38] elaborate on a methodology to extract events using a general-purpose parser and grammar applied to the biomedical domain. To this extent, lexico-syntactic patterns are employed that define the argumentation structures within texts. Hung et al. [13] elaborate on a framework that can be employed for mining the Web for event-based commonsense knowledge by using lexico-syntactic pattern matching and semantic role labeling. A large number of raw sentences that possibly contain target knowledge is collected through Web search engines. Web queries are formulated based on a set of lexico-syntactic patterns. After labeling the semantic roles, i.e., defining the relationships that syntactic arguments have with verbs, knowledge is extracted and stored in a database. A final example of the employment of lexico-syntactic patterns can be found in the work of Xu et al. [37]. Here, the authors envisage the usage of lexico-syntactic patterns in order to learn patterns from texts on prize award events, in the form of bootstrapping-oriented unsupervised machine learning, initialized with lexico-syntactic pattern seeds.

In pattern-based event extraction, concepts that have specific meanings and/or relationships are required, but either they are not available or they are not used due to the lack of pattern expressivity (i.e., in lexico-syntactic patterns). To solve this, lexico-semantic patterns are employed. These patterns are used for various purposes. In an attempt to discover event patterns from stock market bulletins, the authors of [20] analyze tagged corpora by means of gazetteering semantic concepts that are based on a (financial) domain. Cohen et al. [6] employ a concept recognizer on a biological domain in order to extract medical events from corpora, thus taking into account the semantics of domain concepts.

A similar approach is used by Vargas-Vera and Celjuska [35], who propose a framework for event recognition, focusing on Knowledge Media Institute (KMi) news articles. The framework aims for learning and applying lexico-semantic patterns. The extracted information is utilized to populate a knowledge base. Lastly, Capet et al. [3] present a methodology aimed at event extraction for an automated early warning system. The authors employ lexico-semantic patterns for concept matching using dependency chains enhanced using lexicons (word lists), so that concepts are matched whenever syntactically related chains of expressions conveying their constituent concepts occur within the same sentence.

Several advantages stem from the utilization of pattern-based approaches to event extraction. Firstly, pattern-based approaches need less training data than data-driven approaches. Also, it is possible to define powerful expressions by using lexical, syntactical, and semantic elements, and results are easily interpretable and traceable. Patterns are useful when one needs to extract very specific information. However, in order to be able to define patterns that retrieve the correct, desired information, lexical knowledge and possibly also prior domain knowledge is required. Other disadvantages are related to defining and maintaining patterns, as knowledge acquisition is made more difficult (e.g., in costs and consistency) when patterns need to be scaled-up to cover more situations [33] due to the fact that patterns are usually hand-tuned.

2.3 Hybrid Event Extraction

Despite the advantages of both data-driven and knowledge-driven approaches to event extraction, in practice, it is difficult to stay within the boundaries of a single event extraction approach. As both approaches have their disadvantages, combining the two methods could yield the best results. In general, an approach can be viewed as mainly data or knowledge-driven. However, there is an increasing number of researchers that equally combine both approaches, and thus in fact employ hybrid approaches. For instance, it is hard to apply solely pattern-based algorithms successfully, as these algorithms often need for instance bootstrapping or initial clustering, which can be done by means of statistics [29]. Hybrid approaches could emerge when solving the lack of expert knowledge for pattern-based approaches, by applying statistical methods [5]. Also, researchers can combine statistical approaches with (lexical) knowledge, e.g. to prevent unwanted results [28] or to reinforce statistical methods [30].I addition, you can also constrain the learning methods (i.e. data-driven approaches) by using expert knowledge so that a better model is learnt more easily.

In IE literature, many hybrid approaches to text mining are described for extracting events. Most systems are knowledge-driven methods that are aided by data-driven methods, and thus frequently solve the lack of expert knowledge or apply bootstrapping to boost extraction performances, e.g., in terms of precision and recall. For instance, Jungermann and Morik [16] combine lexico-syntactic patterns with conditional random fields (depicted as undirected graphs), in order to extract events from the minutes of plenary sessions of the German parliament. An example of bootstrapping lexical techniques with statistics is given in [29]. Here, the authors bootstrap a weakly supervised pattern learning algorithm with clusters, in order to be able to extract violelence incidents from online news with high precision and recall, as well as storing these in knowledge bases. Chun et al. [4] extract events from biomedical literature by means of lexico-syntactic patterns, combined with term co-occurrences. Finally, aiming for ontology-based fuzzy event extraction for Chinese e-news summarization, the authors of [18] employ a grammar-based statistical method to text mining, i.e., part-of-speech tagging. However, tagging is based on domain knowledge that is stored in ontologies, thus making the event extraction a hybrid process.

In hybrid event extraction systems, due to the usage of data-driven methods, the amount of required data increases, yet typically remains less than is the case with purely data-driven methods. Compared to a knowledge-driven approach, complexity – and hence required expertise – increases due to the combination of multiple techniques. On the other hand, the amount of expert knowledge that is needed for effective and efficient event discovery is generally less than for pattern-based methods, because of the fact that lack of domain knowledge can be compensated by the use of statistical methods. As for the interpretability, attributing results to specific parts of the event extraction is more difficult due to the addition of data-driven methods. Yet, interpretability still benefits from the use of semantics. Disadvantages of hybrid approaches are mostly related to the multidisciplinary aspects of hybrid systems.

3 Discussion

Table 1 provides a summary of the methods discussed, by combining the results from the discussions in Section 2. Per approach elaborated on in this paper, the employed methods and the type of events that are discovered are summarized. Also, the minimum amount of required data and required domain knowledge and expertise are included, as well as the interpretability of the results.

From the results presented in this table, we derive that in terms of data usage, knowledge-driven event extraction methods require the least amount of data (i.e., experiments are performed on a couple of hundreds of documents or sentences). Data-driven methods on the other hand often make use of more than ten thousand documents. Hybrid methods generally report results on a maximum of ten thousand documents. As for interpretability, i.e., the ease with which the (intermediate) results can be translated to a human-understandable format, data-driven methods perform worst. Knowledge-driven methods on the other hand score higher on interpretability. Especially lexico-semantic pattern approaches have a high level of interpretability, as patterns can easily be translated into natural language, while lexico-syntactic patterns require more effort. Finally, when considering the amount of expert domain knowledge and expertise needed for each approach, data-driven methods require less of both than hybrid and knowledge-driven methods.

As a general guideline for selecting a suitable technique for event extraction, based on the results of our survey, we suggest the usage of knowledge-based

Table 1. Overview of the approaches discussed, displaying the method (*Method*) and the type of events that are discovered (*Events*). Also, the amount of required data (*Data*) is depicted, as well as required domain knowledge and expertise (*Know.* and *Exp.*, respectively), and the interpretability of the results (*Int.*). Note that the reported values in the last four columns are lower bounds.

Technique	Approach	Method	Events	Data	Know.	Exp.	Int.
Data	Okamoto et al. [27]	Hierarchical clustering	Local	Med	Low	Low	Low
	Liu et al. [21]	Graphs, clustering	News	High	Low	Low	Low
	Tanev et al. [34]	Clustering	Violent and disaster news	Med	Low	Low	Low
	Lei et al. [19]	Support Vector Machines	News	High	Low	Low	Low
Knowledge	Nishihara et al. [26]	Lexico-Syntactic	Personal experiences	Low	Med	High	Med
	Aone et al. [1]	Lexico-Syntactic	General	Low	High	High	Med
	Yakushiji et al. [38]	Lexico-Syntactic	Biomedical	Low	Med	High	Med
	Hung et al. [13]	Lexico-Syntactic	Commonsense knowledge	Low	Med	High	Med
	Xu et al. [37]	Lexico-Syntactic	Prize award	Low	Med	High	High
	Li et al. [20]	Lexico-Semantic	Financial	Low	High	High	Med
	Cohen et al. [6]	Lexico-Semantic	Biomedical	Med	High	High	High
	Vargas-Vera et al. [35]	Lexico-Semantic	KMi news	Low	High	High	High
	Capet et al. [3]	Lexico-Semantic	Early warning	Low	High	High	High
Hybrid	Jungermann et al. [16]	Lexico-Syntactic, graphs	German parliament	Med	Med	High	Med
	Piskorski et al. [29]	Lexico-Semantic, clustering	Violent news	High	Med	Med	Med
	Chun et al. [4]	Lexico-Syntactic, co-occurences	Biomedical	Med	Med	Med	Med
	Lee et al. [18]	Ontology-based Part-Of-Speech tagging	Chinese news	N/A	Med	Med	Low

techniques for casual users (e.g., students) that prefer an interactive, query-driven approach to event extraction, assuming domain knowledge and expertise to be readily available. Users can easily specify patterns in a language that is close to their own natural language, without being bothered with statistical details and model fine-tuning. On the other hand, users like (academic) researchers would benefit from both hybrid and data-driven approaches, as these are less restricted by, for example, grammars.

4 Conclusions

In this paper, we investigated the main approaches to event extraction from text that are elaborated on in the current body of literature. Overall, data-driven methods require many data and little domain knowledge and expertise, while having a low interpretability. Conversely, for knowledge-based event extraction little data is required, but domain knowledge and expertise is needed. These approaches generally offer a higher traceability of the results. Finally, hybrid approaches seem to be a compromise between data and knowledge-driven approaches, requiring a medium amount of data and domain knowledge and offering medium interpretability. However, it should be noted that the amount of expertise needed is high, due to the fact that multiple techniques are combined. As a guideline, we advise knowledge-driven techniques for casual and novice users, whereas data-driven are more suitable for advanced users.

References

1. Aone, C., Ramos-Santacruz, M.: REES: A Large-Scale Relation and Event Extraction System. In: 6th Applied Natural Language Processing Conference (ANLP 2000). pp. 76–83. Association for Computational Linguistics (2000)
2. Borsje, J., Hogenboom, F., Frasincar, F.: Semi-Automatic Financial Events Discovery Based on Lexico-Semantic Patterns. International Journal of Web Engineering and Technology 6(2), 115–140 (2010)
3. Capet, P., Delavallade, T., Nakamura, T., Sandor, A., Tarsitano, C., Voyatzi, S.: Intelligent Information Processing IV, IFIP International Federation for Information Processing, vol. 288, chap. A Risk Assessment System with Automatic Extraction of Event Types, pp. 220–229. Springer Boston (2008)
4. Chun, H.W., Hwang, Y.S., Rim, H.C.: Unsupervised Event Extraction from Biomedical Literature Using Co-occurrence Information and Basic Patterns. In: 1st International Joint Conference on Natural Language Processing (IJCNLP 2004). Lecture Notes in Computer Science, vol. 3248, pp. 777–786. Springer-Verlag Berlin Heidelberg (2004)
5. Cimiano, P., Staab, S.: Learning by Googling. SIGKDD Explorations Newsletter 6(2), 24–33 (2004)
6. Cohen, K.B., Verspoor, K., Johnson, H.L., Roeder, C., Ogren, P.V., Baumgartner, Jr., W.A., White, E., Tipney, H., Hunter, L.: High-Precision Biological Event Extraction with a Concept Recognizer. In: Workshop on BioNLP: Shared Task collocated with the NAACL-HLT 2009 Meeting. pp. 50–58. Association for Computational Linguistics (2009)

7. Cunningham, H.: GATE, a General Architecture for Text Engineering. Computers and the Humanities 36(2), 223–254 (2002)
8. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: 40th Anniversary Meeting of the Association for Computational Linguistics (ACL 2002). pp. 168–175. Association for Computational Linguistics (2002)
9. Ferrucci, D., Lally, A.: UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. Natural Language Engineering 10(3–4), 327–348 (2004)
10. Frasincar, F., Borsje, J., Levering, L.: A Semantic Web-Based Approach for Building Personalized News Services. International Journal of E-Business Research 5(3), 35–53 (2009)
11. Hearst, M.A.: Automatic Acquisition of Hyponyms from Large Text Corpora. In: 14th Conference on Computational Linguistics (COLING 1992). vol. 2, pp. 539–545 (1992)
12. Hearst, M.A.: WordNet: An Electronic Lexical Database and Some of its Applications, chap. Automated Discovery of WordNet Relations, pp. 131–151. MIT Press (1998)
13. Hung, S.H., Lin, C.H., Hong, J.S.: Web Mining for Event-Based Commonsense Knowledge Using Lexico-Syntactic Pattern Matching and Semantic Role Labeling. Expert Systems with Applications 37(1), 341–347 (2010)
14. Ikenberry, D.L., Ramnath, S.: Underreaction to Self-selected News Events: The Case of Stock Splits. Review of Financial Studies 15(2), 489–526 (2002)
15. Ikonomakis, M., Kotsiantis, S., Tampakas, V.: Text Classification Using Machine Learning Techniques. WSEAS Transactions on Computers 4(8), 966–974 (2005)
16. Jungermann, F., Morik, K.: Enhanced Services for Targeted Information Retrieval by Event Extraction and Data Mining. In: 13th International Conference on Natural Language and Information Systems: Applications of Natural Language to Information Systems (NLDB 2008). Lecture Notes in Computer Science, vol. 5039, pp. 335–336. Springer-Verlag Berlin Heidelberg (2008)
17. Kamijo, S., Matsushita, Y., Ikeuchi, K., Sakauchi, M.: Traffic monitoring and accident detection at intersections. IEEE Transactions on Intelligent Transportation Systems 1(2), 108–118 (2000)
18. Lee, C.S., Chen, Y.J., Jian, Z.W.: Ontology-Based Fuzzy Event Extraction Agent for Chinese E-News Summarization. Expert Systems with Applications 25(3), 431–447 (2003)
19. Lei, Z., Wu, L.D., Zhang, Y., Liu, Y.C.: A System for Detecting and Tracking Internet News Event. In: 6th Pacific-Rim Conference on Multimedia (PCM 2005). Lecture Notes in Computer Science, vol. 3767, pp. 754–764. Springer-Verlag Berlin Heidelberg (2005)
20. Li, F., Sheng, H., Zhang, D.: Event Pattern Discovery from the Stock Market Bulletin. In: 5th International Conference on Discovery Science (DS 2002). Lecture Notes in Computer Science, vol. 2534, pp. 35–49. Springer-Verlag Berlin Heidelberg (2002)
21. Liu, M., Liu, Y., Xiang, L., Chen, X., Yang, Q.: Extracting Key Entities and Significant Events from Online Daily News. In: 9th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2008). Lecture Notes in Computer Science, vol. 5326, pp. 201–209. Springer-Verlag Berlin Heidelberg (2008)
22. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press, 1st edn. (1999)

23. Michaely, R., Thaler, R.H., Womack, K.L.: Price Reactions to Dividend Initiations and Omissions: Overreaction or Drift? Journal of Finance 50(2), 573–608 (1995)
24. Mitchell, M.L., Mulherin, J.H.: The Impact of Public Information on the Stock Market. Journal of Finance 49(3), 923–950 (1994)
25. Moreau, N.: Best Practices in Language Resources for Multilingual Information Access. Tech. rep., TrebleCLEF Consortium (2009), From: `http://www.trebleclef.eu/getfile.php?id=255`
26. Nishihara, Y., Sato, K., Sunayama, W.: Event Extraction and Visualization for Obtaining Personal Experiences from Blogs. In: Symposium on Human Interface 2009 on Human Interface and the Management of Information. Information and Interaction. Part II. Lecture Notes in Computer Science, vol. 5618, pp. 315–324. Springer-Verlag Berlin Heidelberg (2009)
27. Okamoto, M., Kikuchi, M.: Discovering Volatile Events in Your Neighborhood: Local-Area Topic Extraction from Blog Entries. In: 5th Asia Information Retrieval Symposium (AIRS 2009). Lecture Notes in Computer Science, vol. 5839, pp. 181–192. Springer-Verlag Berlin Heidelberg (2009)
28. Pakhomov, S.: Semi-Supervised Maximum Entropy Based Approach to Acronym and Abbreviation Normalization in Medical Texts. In: 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002). pp. 160–167. Association for Computational Linguistics (2002)
29. Piskorski, J., Tanev, H., Wennerberg, P.O.: Extracting Violent Events From On-Line News for Ontology Population. In: 10th International Conference on Business Information Systems (BIS 2007). Lecture Notes in Computer Science, vol. 4439, pp. 287–300. Springer-Verlag Berlin Heidelberg (2007)
30. Punyakanok, V., Roth, D., Yih, W.: The Importance of Syntactic Parsing and Inference in Semantic Role Labeling. Computational Linguistics 34(2), 257–287 (2008)
31. Rosen, R.J.: Merger Momentum and Investor Sentiment: The Stock Market Reaction to Merger Announcements. Journal of Business 79(2), 987–1017 (2006)
32. Schouten, K., Ruijgrok, P., Borsje, J., Frasincar, F., Levering, L., Hogenboom, F.: A Semantic Web-Based Approach for Personalizing News. In: 25th Symposium On Applied Computing (SAC 2010). pp. 854–861. ACM (2010)
33. Su, K.Y., Chiang, T.H., Chang, J.S.: An Overview of Corpus-Based Statistics-Oriented (CBSO) Techniques for Natural Language Processing. Computational Linguistics and Chinese Language Processing 1(1), 101–157 (1996)
34. Tanev, H., Piskorski, J., Atkinson, M.: Real-Time News Event Extraction for Global Crisis Monitoring. In: 13th International Conference on Natural Language and Information Systems: Applications of Natural Language to Information Systems (NLDB 2008). Lecture Notes in Computer Science, vol. 5039, pp. 207–218. Springer-Verlag Berlin Heidelberg (2008)
35. Vargas-Vera, M., Celjuska, D.: Event Recognition on News Stories and Semi-Automatic Population of an Ontology. In: 3rd IEEE/WIC/ACM International Conference on Web Intelligence (WI 2004). pp. 615–618 (2004)
36. Wei, C.P., Lee, Y.H.: Event detection from Online News Documents for Supporting Environmental Scanning. Decision Support Systems 36(4), 385–401 (2004)
37. Xu, F., Uszkoreit, H., Li, H.: Automatic Event and Relation Detection with Seeds of Varying Complexity. In: AAAI Workshop on Event Extraction and Synthesis (2006)
38. Yakushiji, A., Tateisi, Y., Miyao, Y.: Event Extraction from Biomedical Papers using a Full Parser. In: 6th Pacific Symposium on Biocomputing (PSB 2001). pp. 408–419 (2001)