

# Event characterization for information extraction from business-based technical documents

**Bell Manrique Losada**

Universidad de Medellín  
Cra. 87 30-65 Belén  
Medellín, AQ, Colombia  
bmanrique@udem.edu.co

**Carlos M. Zapata Jaramillo**

Universidad Nacional de Colombia  
Cra. 80 65-223 Robledo  
Medellín, AQ, Colombia  
cmzapata@unal.edu.co

## Abstract

Event identification plays a crucial role in several natural language processing applications such as information extraction, question answering, and text analysis. In this paper, we describe a novel approach for analyzing events, their distribution, and the event mentions from a corpus of unlabeled business-based technical documents—a specific genre. In order to infer such mentions, we analyze the subject-verb-object structure for semi-automatically extracting several lexical, syntactic, and semantic features for each event mention from the corpus. Extracting event mentions allows us to cast grouping together the mentions with same features and propose properties leading to the differences of the specific genre. The obtained results are used for supporting an event-centered processing level, from an automated machine for processing texts.

## 1 Introduction

Information extraction (IE) is a process for extracting structured information from unstructured texts (Sangeetha *et al.*, 2010). Event identification plays a crucial role in IE and other natural language processing applications (*e.g.*, question answering, text analysis, etc.). Identification of event structures can exploit cross-document techniques. A special attention is given to the recognition of events from heterogeneous document sources, stemming from several genres and domains (Petrenz and Webber, 2011).

According to Pivovarova *et al.* (2013), in the context of IE, events represent real-world facts and they should be extracted from plain text. By the

unique nature of events, they receive in-depth attention in current research, by trying to identify what events are mentioned within texts and how they are related semantically (Do *et al.*, 2011).

In this paper we propose an unsupervised approach for identifying events, from unannotated sentences of business-based technical documents contained in a training corpus. We base our approach on text expressions referring to real-world events—also called event mentions—for identifying events (Bejan & Harabagiu, 2013) from a set of clusters.

We define training documents based on lexical chains by holding the set of semantically related words of given sentences. WordNet lexicon was used for constructing lexical chains with the event mentions. A set of features and properties for each event has been identified in order to obtain a characterization of the specific genre concerned to the technical document.

The preliminary results, in terms of events features and properties, are based on a processing level in an automated system for processing texts. Then, such event-centered processing level is the basis for identifying the organizational domain knowledge and some business information as the first instances of the requirement elicitation process.

The remainder sections of this paper are organized as follows: in Section 2 we describe the related work in the field of information extraction and event extraction. In Section 3 we present our approach to event identification for extracting information from the genre *business-based technical documents*. Finally, in Section 4 we draw conclusions and we outline future work.

## 2 Overview and related work

This Section provides a short overview and related work of the most relevant research in terms of event analysis, information extraction, and event extraction.

### 2.1 Events

Events represent real-world facts. Also, they may have several relationships to such facts, and different sources may have contradictory views on the facts (Saurí & Pustejovsky, 2012). Thus, the structure and content of an event is influenced by both the structure of the specific real-world fact and the properties of the surrounding text. The role of the events in a text depends directly on the context, real-world domain, or scenario in which the text is used. In this sense, events are representations of facts and also linguistic units. Whereby, and according to Pivovarova *et al.* (2013), in the analysis and research of events should be considered the particular language, genre, scenario and medium of the text—*i.e.*, events should be analyzed in the context of particular corpora.

Our motivation yields on the event study in practice, looking for identifying domain-specific characteristics of events in a business-based corpus. We hope this preliminary study of the corpus can be used in the same or greater depth of linguistic analysis by a language processing system or an IE system.

### 2.2 Basics of lexical-semantic analysis

WordNet is a lexical semantic resource which defines word senses by using methods for grouping senses of the same word and thus producing coarser word sense groupings (Fellbaum, 1998). For the aim of this work and looking for analyzing events, we consider the syntactic categories of verbs. Verbs form language-specific structures in the WordNet ontology and they are included in the category of 2nd Order Entity. According to Vossen (2002), such a category comprises entities referring to any situation—being static or dynamic—which cannot be grasped, heard, seen, or felt as an independent physical thing. These situations can occur or take place in a time or place/space, rather than exist (*e.g.* happen, cause, occur, apply, etc.). Also, they are related to: i) verbs or events denoting

nouns, and ii) events, processes, states-of-affairs, or situations located in time. Verbs in this category can be further subrogated according to the physical entities involved in the following subcategories:

**Process.** This category implies all physical entities, *i.e.*, those located in space-time. Entities related to objects and processes are involved in it. Verbs in this category are mostly related to processes since they are things that ‘happen’ and have ‘temporal parts/stages’. A process can be considered as a set of denotations related to dual object process, intentional process, motion, internal change, shape, or change.

**Situation Type.** Refers to a situation—event or set of events, featured as a conceptual unit—happening over time. This kind of verbs is represented in terms of the event-structure or the predicate properties.

**Conceptual Domain.** EuroWordNet is a multilingual database for multiple languages containing 200 domain labels organized in a hierarchical structure for grouping the words in categories based on a domain hierarchy. Semantic domains are knowledge areas—*e.g.* economy or politics—used to describe texts according to general subjects characterized by domain specific lexica. The domain hierarchy is represented as an ontology which comprises conceptual levels for each language. The levels of the domain hierarchy are called basic domains.

### 2.3 Language processing techniques

Several language processing techniques centered in events have been used in areas such as text-mining and information extraction. They have been applied by considering many kinds of documents, *e.g.*, technical documents, patents, and software requirement documents, as follows.

Cascini *et al.* (2004) present a functional analysis of patents and their implementation in the PAT-Analyzer tool. They use techniques based on the extraction of the interactions from the entities described in the document and expressed as subject-action-object triples, by using a suitable syntactic parser. Rösner *et al.* (1997) generate multilingual documents from knowledge bases by using automated techniques. The resulting documents can be represented in an interchangeable way centered in events.

## 2.4 Information Extraction and Text Analytics

*Information Extraction* includes techniques for extracting any kind of information from texts. Relation extraction techniques require the identification of significant entities and relationships between entities and significant properties of entities (Grimes, 2008). The goal of IE is storing the extracted entities and relationships in a database—structured information. The *prototypical document extraction* relies on the identification of frequent sequences of terms in the documents, and uses language processing techniques, such as POS tagging and term extraction, for pre-processing the textual data (Rajman & Besancon, 1997). Such a technique can be considered as an automated, generalized indexing procedure for extracting linguistically significant structures from documents.

According to Wilcock (2009), *Text Analytics* (TA) refers to a subfield of information technology dealing with applications, systems, and services for doing some kind of text analytics as a way to extract information from them. Several techniques for TA have been developed, among them: named entity recognition, co-reference resolution, information extraction, chunking, semantic role labeling, text mining, and semantic search.

The state-of-the-art review present several approaches in the previous areas for studying events, as follows.

Meth *et al.* (2012) propose an automated and knowledge-based support system for eliciting activities and process, in the context of knowledge engineering. RARE project (Cybulski & Reed, 1998) is focused on parsing texts based on a semantic network assisted by a thesaurus; they combine NLP with faceted classification for identifying and analyzing needs and expectations of stakeholders. Hahn *et al.* (1996) develop a methodology for knowledge acquisition and concept learning from texts written in German. The method relies on a quality-based model for reasoning on terminology, by using concepts from NLP.

Focused on goal identification we found several approaches (Dardenne *et al.*, 1993; Darimont *et al.*, 2005; Giorgini *et al.*, 2005). Such goals describe desired states or actions performed by actors regardless of specific consideration for normative positions (*e.g.*, permissions, recommendations, and obligations). Young and Antón (2010) propose the

analysis of the commitments, privileges, and rights conveyed within online policy documents.

## 2.5 Event extraction

The event extraction has been approached by several authors as we present in the following paragraphs.

Huttunen *et al.* (2002a) propose linguistic cues for identifying the overlapping or partial events including specific lexical items, locative and temporal expressions, and usage of ellipsis and anaphora. Grishman (2012) emphasizes in unsupervised event extraction by using extensive linguistic analysis. Do *et al.* (2011) develop a minimally supervised approach, based on focused distributional similarity methods and discourse connectives, for identifying causality relations between events in context. Sun *et al.* (2007) are focused on detecting causality between search query pairs in temporal query logs.

Riaz and Girju (2010) propose cluster sentences into topic-specific scenarios, and then focus on identifying causal relations between events and building a dataset of causal text spans headed by a verb. Etzion and Niblett (2010) work with event processing and present a software system including specific logic to filter, transform, or detect patterns in events as they occur. The event analysis in specific genres has been approached as follows:

- Beamer and Girju (2009) work on detecting causal relations among verbs in a corpus of screen plays, limited to consecutive or adjacent verb pairs.
- Szarvas *et al.* (2012) study the linguistic cues of events in three genres: news, scientific papers, and Wikipedia articles. They demonstrate significant differences in lexical usage across the genres by using syntactic cues.
- Pivovarova *et al.* (2013) propose the event analysis for generating particular statistics and capturing the scenario-specific characteristics of event representation in a particular corpus.
- The PULS<sup>1</sup> system is based on the event structure for discovering, aggregating, verifying, and visualizing events in various scenarios.

Finally, relevant proposals for event extraction have been developed: Chambers and Jurafsky (2011) propose a template-based IE algorithm for

---

<sup>1</sup> <http://puls.cs.helsinki.fi>

learning sets of related events and semantic roles from an unlabeled corpus; Kasch and Oates (2010) define script learning and narrative schemas to capture knowledge from unlabeled text. Scripts are sets of related event words and semantic roles learned by linking syntactic functions with corefering arguments; Benson *et al.* (2011) propose a method for discovering event records from social media feeds. Such a method operates on a noisy feed of data and extracts canonical records of events by aggregating information across multiple messages.

### 3 Approach for event analysis

#### 3.1 Corpus and Analysis Tools

The corpus definition starts by collecting possible technical documents circulating on the web related to the genre *business domain*. We have not so many restrictions by selecting the texts for building the corpus, because the focus is getting as many samples as possible, but not the entire track rolling stock. We collect and analyze a set of documents from such a domain in different subject fields (*e.g.* medicine, forestry, and laboratory). The corpus used as the basis for this preliminary study comprises 50 English-written documents with independence of its variety. Assuming the population is evenly distributed, we selected a sample of 32 documents, corresponding to 64% of the total corpus population—the minimum percentage statistically random, calculated with proportions Z test. The variety of subject fields is important to the analysis of the events identified in the corpus.

The training documents belong to the ‘Standard Operating Procedure (SOP) category. All the documents sum 167,905 tokens and 9,252 word types. The initial exploration of this experimental corpus was supported by AntConc 3.3.5w® (Anthony, 2009) and TermoStatWeb™ (Drouin, 2003). AntConc was used to manually and systematically find frequent expressions and select their contexts, and TermoStatWeb™ was used to list most frequent verbs, for analyzing its organization in the texts.

#### 3.2 Analysis approach

This analysis is based on the semantic behavior of the events, under the premise the analysis of mean-

ings or senses of the verbs should be closely linked to the analysis of events and terms used in a context. This event-centered analysis is approached from the point of view of the possible meanings suggested by the Multilingual Central Repository (MCR)<sup>2</sup>.

Based on the training corpus (SOP), we identify the set of most frequent verbs. Prioritized verbs are classified by categories, according to Vossen (2002). Then, we use the types of verb in order to identify patterns. Such patterns will be the basis of rules for inferring and extracting organizational relationships from business-based information. In this way, we guide the analysis to all situations concerning the verb regarding its usage in the SOPs. Based on an incremental method, we performed this step-by-step analysis as follows:

**Review.** In this phase we look for identifying the verbs in the relevant sections of the documents, according to the rhetorical organization units defined by Manrique (2015). For the sake of identifying verbs, we first prioritize the most used verbs in the SOPs according to the occurrence frequency. Analysis of the occurrence frequency of verbs in the corpus was supported by corpus analysis tools. We selected the first 58 verbs corresponding to the interval of hits [501–72], with 501 the highest frequency and 72 the lower occurrence. In Table 1 we present the 10 most used verbs in the corpus.

Term	Frequency
use	501
include	458
provide	355
follow	314
require	314
ensure	226
submit	175
approve	166
prepare	156
identify	143
involve	143
perform	136
describe	133

Table 1. Sample of prioritized verbs

<sup>2</sup> <http://adimen.si.ehu.es/web/MCR>

**Verb feature identification.** The semantic features of each prioritized verb are addressed, and the discriminant features for relevant and irrelevant verbs, according to some categories—conceptual, functional, process-based, and situational—are identified. The classification of verbs is based on the meanings reported by WordNet 3.0 by using the Interlingual Index (ILI 3.0) included into the EuroWordNet Interface (WEI consult mode)<sup>3</sup>. For each prioritized verb, we define the categories of classification according to the classes defined in section 2.2, by following an iterative and sequential process:

i) For each verb, checking the meanings reported in WEI. For the first four meanings reported:

- Assign an occurrence indicator (value) for each associated conceptual category.
- Assign a value for each functional category the verb has.
- Assign values to situational category (unique value for situation type) and multiple value for the situation component.

ii) Generating the sum of all values for each category

iii) Identifying which categories correspond to the highest sum of the corresponding analysis. The complete verb classification was presented by Manrique (2015), as a result of the previous analysis. In Table 2 a sample of such results is presented, where: column 1 is de list of prioritized verbs; columns 2 to 6 correspond to the conceptual classification category of verb (based on WordNet 3.0 and ILI). The number appearing in each cell verb/category corresponds to the frequency of verb is taking such a category. We generate the sum of all values by category and identify the categories corresponding with the highest sum.

iv) Analyzing and presenting results. Based on the defined classification and categorization of verbs, we make an analysis and identify findings in terms of features and properties capturing the particularities of the specific genre.

<sup>3</sup> <http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl>

### 3.3 Characterization of events from business-based technical documents

As a result of the previous process and according to the each analyzed and prioritized category, we identify the following findings, characterizing the genre:

**Conceptual Category.** The conceptual domains are based on a relation of specificity. We identified the events are not assigned to a particular conceptual category, due to their nature. Unlike nouns which somehow can be grouped by domains, the events/verbs can be used in similar senses by several different domains. For instance we can find that ‘*apply*’ is used in a specific domain like ‘*Medicine*’ for saying ‘*a nurse is applying medication to a patient,*’ or in a general domain involving an intentional process for saying ‘*The operator works applying rules.*’

Verb	Conceptual classification of verb (Category)												
	facto tum	social	free-time	applied science	humanities	pure-sciences	social	Possession	change	cognition	stative	communication	creation
assign							1						
use	3	1					1						
include	3						1		1				
provide	2	1	1				1					1	
follow	4							2					
require	4											2	
review	2	1			1					3		1	
process	1	1		2		1	2	1	1				
ensure	2									1		1	
request	3											3	
submit	4							1				2	
work	3	1					3						
approve	2									1		1	
prepare	3		1						1				3

Table 2. Sample of prioritized verbs

As we show in the results of classification, the prioritized verbs are mostly used in any domain, for example the one appearing under the label *Factotum*, which is assigned when none of the labels were assigned. When verb is not labeled as *factotum*, the second mostly used conceptual category is *social label*.

**Functional category.** According to the tracks of the analysis, most of verbs are marked as an *intentional process* (general), whose intention is no longer identified (e.g. attaching, comparing, substituting, and separating). Generally speaking, an *intentional process* is deliberately set in motion by a Cognitive Agent, i.e. it is a human action, act, or activity of a thing for accomplishing or achieving a work.

The second most frequent functional category is *social interaction* as a kind of intentional process involving interactions between Cognitive-Agents. This category relates a social relation, an interaction, or a socially accepted situation.

**Situation Category.** The *situation type* for most of the verbs is *dynamic*. Such verbs are related to the situations implying either a transition from one state to another or a continuous transition perceived as an unbounded process (e.g. event, act, action, become, happen, take place, process, habit, change, and activity). No change in their properties or relation is involved by the verb.

**Dynamic situations.** More than half of verbs occur with *bounded event*, when they are implied with a specific transition from one situation to another, which is bounded in time and directed to a result (e.g., *to implement, to remove, to develop, etc.*). Regarding to the *situation component*, the results show the main semantic components characterizing the situation are the following:

- *Cause.* Component of situations involving causation. This component is coherent with the *situation type* since the *causation* is always combined with *dynamic* and it can take several forms. Such forms depend on the grade of intervention of an agent. The form with the higher frequency is *agentive* which can be related to a controlling agent who intentionally tries to achieve some change. The agentive situations imply a controlling agent causing a dynamic change (e.g. *to implement, to write, to record, etc.*).
- *Purpose.* Abstract components reflecting the intentionality of acts and activities. Situations intended to have some effect are implied. As the previous situation component, this one reflects consistency with the context as applied to *dynamic situations*. Also, this component strongly correlates to *agentive* and *cause*, clustering mainly hu-

man acts and activities. *Situation Components* such as *usage, social, and communication* often combine with *purpose*.

- *Communication.* Component of situations involving communication (e.g. designate, request, describe, issue, etc.). Communication verbs are often speech-acts (*bounded events*) or denote more global communicative activities (*unbounded events*). Also, they include different phases of the communication referring to causation of communication effects (e.g., *to explain or to show*) or creation of a meaningful representation (e.g., *to write or to draw*).
- *Physical.* Component of situations involving perceptual and measurable properties of objects (e.g., *to shape, to prepare, to describe, etc.*); or dynamic changes and perceptions of its physical properties (e.g., *to monitor, to collect, to copy, to notice, etc.*).

Based on the previous characterization (features and properties) and the prioritized verbs, we finally develop a dependency parsing. For the parsing process we use the Freeling dependency parser<sup>4</sup>. The goals of such parsing are:

- Defining patterns of occurrence of the identified verbs and defining a set of semantic/dependency rules for transforming each pattern to a controlled language structure.
- Defining a script for preprocessing the SOPs, trying to extract simple sentences for the parser to maximize its performance.
- Processing the evaluation corpus with a dependency parser.
- Evaluating the extracted relations and the findings.

According to the event characterization and feature identification resulted from the dependency parsing, we propose a set of semantic rules for transforming each feature into a controlled language. We used the UN-Lencep (named by its Spanish acronym for ‘Universidad Nacional de Colombia—Lenguaje Controlado para la Especificación de Esquemas Preconceptuales’), as an intermediate representation between natural language and conceptual schemas for software engineering.

---

<sup>4</sup> <http://nlp.lsi.upc.edu/freeling/>

We present the rules defined for such mapping in Manrique & Zapata (2013). Each mapping rule is assigned to one category expressed in terms of the pattern in the SOP and the expression in UNLencep generated. The pattern composing each defined rule considers attributes relating the tags (*e.g.*, syntactic or semantic tag—synt—, function tag—func—, etc.) assigned by the dependency parser.

By means of the basic interface of the parser library, we analyzed the text files of the corpus from the command line. We extract the relations matching the semantic rules from the parser. Based on them, we performed a preliminary evaluation in terms of the useful relations being extracted, the number of relevant extracted relations with the necessary components for measuring precision and recall. According to the results, we could identify the potential of the parsing, the quality of the defined rules, and the aspects improved by the text preprocessing.

## 4 Conclusions

This study aims at characterizing SOPs by revealing key features and properties of events used in an English corpus belonging to the business genre. We proposed an approach for analyzing events from a training corpus, which can be processed as input for further knowledge engineering processes. The appropriateness of JSDs in requirements elicitation was verified with this study.

We analyze the structure of training text for semi-automatically extracting several features for each event mention from the corpus. Extracting a rich set of features allows us propose properties capturing the differences of this specific genre. We contribute to the research about the identification of events from heterogeneous document sources stemming from different genres and domains

Our proposal is focused on the events study in the practice, for identifying domain-specific characteristics of them in a business-based corpus. This is a preliminary study which we expect can be used in the same or greater depth of linguistic analysis by language processing systems or IE systems.

We are testing the performance of the rules derived from this event analysis approach in NAHUAL, our functional prototype of a software system for processing texts.

As future work, we expect to increase the number of documents in the corpus and refine the study of event features. Statistical measures can be also considered as a way to support the presented event analysis and the event representation in this particular corpus, as suggest Pivovarova *et al.* (2013). The automated event extraction in the frame of knowledge acquisition from business-based documents is also our interest.

Likewise, given the importance of the event structure, the supervised event causality identification and the causal relations analysis seems to be a promising approach in the current research.

## Acknowledgments

This work has been partly founded by the Vicerrectoría de Investigación from both the Universidad de Medellín and the Universidad Nacional de Colombia, under the following projects: “Método de transformación de lenguaje natural a lenguaje controlado para la obtención de requisitos, a partir de documentación técnica”; “Especificación de un lenguaje controlado de Dominio Específico: Fundamentos Lingüísticos y bases de Transformación desde Documentos Técnicos Corporativos en Lenguaje Natural”. Also, it was supported in part by the COLCIENCIAS, under the doctoral program “Generación del Bicentenario, Colombia 2010.”

## References

- Anthony, L. 2009. *Issues in the design and development of software tools for corpus studies: The case for collaboration*. Contemporary corpus linguistics, London: P. Baker Ed.: 87-104.
- Bejan, C.A. and Harabagiu, S. 2013. Unsupervised Event Coreference Resolution. *Computational Linguistics*, 40: 2.
- Benson, E., Haghghi, A. and Barzilay, R. 2011. Event discovery in social media feeds. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, ACL, Stroudsburg, PA, USA, 1: 389-398.
- Cascini, G. Fantechi, A. and Spinicci E. 2004. Natural Language Processing of Patents and Technical Documentation. *Lecture Notes in Computer Science*, 3163:508-520.
- Chambers, N. and Jurafsky, D. 2011. Template-Based Information Extraction without the Templates, ACL-2011, Portland, OR.

- Cybulski, J. and Reed, K. 1998. Requirements Classification and Reuse: Crossing domains boundaries. *6th Intl. Conf. on Software Reuse (ICSR'2000)*. Vienna, Austria. Springer.
- Dardenne, A., van Lamsweerde, A. and Fickas, S. 1993. Goal-directed Requirements Acquisition. *Science of Comp. Prog.*, 20:3-50.
- Darimont, R., Delor, E., Massonet, P., van Lamsweerde, A. 2005. GRAIL/KAOS: An Environment for Goal-driven Requirements Engineering. *IEEE 19th International Conference on Software Engineering*, pp. 612-613.
- Do, Q.X., Chan, Y.S., and Roth, D. 2011. Minimally Supervised Event Causality Identification. *EMNLP'2011*.
- Drouin, P. 2003. *TermoStat Web 3.0*. Désormais utilisable qu'après enregistrement. Available in: [http://olst.ling.umontreal.ca/~drouinp/termostat\\_web/](http://olst.ling.umontreal.ca/~drouinp/termostat_web/)
- Etzion O., Niblett P. 2010. *Event Processing in Action*. Manning Publications, Co. NY, USA.
- Fellbaum, C. 1998. *WordNet*. An Electronic Lexical Database. The MIT Press. NY, USA.
- Giorgini, P., Massacci, F., Mylopoulos, J., Zannone, N. 2005. Modeling Security Requirements through Ownership, Permission and Delegation. *13th IEEE International Requirements Engineering Conference*, pp. 167-176.
- Grimes, S. 2008. Text technologies in the mainstream: Text analytics solutions, applications and trends. Available in: <http://altaplana.com>
- Grishman, R. 2012. Structural linguistics and unsupervised information extraction. *Automatic Knowledge Base Construction and Web-scale Knowledge Extraction (AKBC-WEKEX 2012)*, pp. 57-61.
- Hahn, U. Klenner, M. and Schnattinger, K. 1996. A Quality-Based Terminological Reasoning Model for Text Knowledge Acquisition. *Advances in Knowledge Acquisition*. Shadbolt, O'Hara & Schreiber (Eds.). Springer-Verlag, Berlin.
- Huttunen, S., Yangarber, R. and Grishman, R. 2002. Diversity of scenarios in information extraction. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Spain.
- Manrique, B. 2015. A formalization for mapping discourses from business-based technical documents into controlled language texts for requirement elicitation. Ph.D. Thesis. Universidad Nacional de Colombia, Medellín.
- Meth, H., Li, Y., Maedche, A. Mueller, B. 2012. Advancing task elicitation systems—An experimental evaluation of design Principles. *Proceeding of 33 International Conference on Information Systems*, pp. 54-68.
- Petrenz, P. and Webber, B. 2011. Stable classification of text genres. *Computational Linguistics*, 37(2):385-393.
- Pivovarova, L., Huttunen, S. and Yangarber, R. 2013. Event representation across genre. *Proceedings of the 1st Workshop on EVENTS: Definition, Detection, Co-reference, and Representation*, pp. 29-37.
- Rajman, M. and Besancon, B. 1997. *Text Mining: Natural Language techniques and Text Mining applications*. IFIP 1997. Chapman & Hall. NY, USA.
- Riaz, M. and Girju, R. 2010. Another look at causality: Discovering scenario-specific contingency relationships with no supervision. *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing*, pp. 361-368.
- Rösner, D., Grote, B., Hartmann, K. and Höfling, B. 1997. From Natural Language Documents to Shareable Product Knowledge: A Knowledge Engineering Approach. *Journal of Universal Computer Science*. 3(8): 955-987.
- Sangeetha S, R.S. Thakur, and Arock, M. 2010. *Event detection using lexical chain*. Hrafn Loftsson, Eiríkur Rögnvaldson, Sigrun Helgadóttir (eds.). In: *IceTAL 2010*, LNAI 6233: 314-319.
- Saurí, R. and Pustejovsky, J. 2012. Are you sure that this happened? Assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261-299.
- Sun, Y., Liu, N., Xie, K., Yan, S., Zhang, B. and Chen, Z. 2007. Causal relation of queries from temporal logs. *Proceedings of the 16th international conference on World Wide Web*, NY, USA.
- Szarvas, G. Vincze, V., Farkas, R., Mófra, G. and Gurevych, I. 2012. Cross-genre and cross-domain detection of semantic uncertainty. *Computational Linguistics*, 38(2):335-367.
- Vossen, P. 2002. *EuroWordNet General Document*. Version 3. Piek Vossen (ed.). University of Amsterdam. Amsterdam, Holanda.
- Wilcock, G. 2009. *Introduction of linguistics annotation and texts analytics*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool (eds). Toronto, CA.
- Young, J. D. and Antón, A. I. 2010. A Method for Identifying Software Requirements Based on Policy Commitments. *18th IEEE International Requirements Engineering Conference*. Available in: <http://ieeexplore.ieee.org.ezproxy.unal.edu.co/stamp/stamp.jsp?tp=&arnumber=5636634>