



Ingeniare. Revista Chilena de Ingeniería

ISSN: 0718-3291

facing@uta.cl

Universidad de Tarapacá

Chile

Lezcano R., Luis Alfonso; Guzmán L., Jaime Alberto; Gómez A., Sebastián Alonso  
Extraction of goals and their classification in the KAOS model using natural language processing  
Ingeniare. Revista Chilena de Ingeniería, vol. 23, núm. 1, enero, 2015, pp. 59-66  
Universidad de Tarapacá  
Arica, Chile

Available in: <http://www.redalyc.org/articulo.oa?id=77233740007>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System  
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal  
Non-profit academic project, developed under the open access initiative

## Extraction of goals and their classification in the KAOS model using natural language processing

### *Extracción de objetivos y su clasificación en el modelo de KAOS a partir del procesamiento del lenguaje natural*

Luis Alfonso Lezcano R.<sup>1</sup>      Jaime Alberto Guzmán L.<sup>1</sup>      Sebastián Alonso Gómez A.<sup>1</sup>

Recibido 15 de octubre de 2013, aceptado 19 de junio de 2014

*Received: October 15, 2013      Accepted: June 19, 2014*

#### ABSTRACT

The KAOS (Knowledge Acquisition in Automated Specification) goals diagram is one of the most important diagrams during software requirements elicitation, that is to say, the first phase of a program's life cycle, since it helps stakeholders (users) to understand the importance of future software. In the studies in order to identify the traceability that should exist between natural language and the goals that have been identified to avoid inaccuracies between them was not possible. This paper presents a method for extracting and classifying goals in KAOS approach through the processing of textual requirements in Spanish. In this method, the phrases found in a text are broken down and a morphological and syntactical analysis is carried out for each phrase, using the classification of Spanish verbs as reference. Furthermore, morphosyntactic structures are defined which allow the goals to be typified and classified, based on the following types: (i) Maintain; (ii) Achieve; (iii) Cease and (iv) Avoid. This classification aims to list all of the goals and to represent them according to the KAOS goal model. This method will serve as a starting point for identifying the other components that make up the KAOS goal diagram and the semiautomatic rendering of said diagram.

Keywords: KAOS methodology, KAOS goal diagram, goal identification, requirement analysis, natural language processing.

#### RESUMEN

*El diagrama de objetivos de KAOS (Knowledge Acquisition in Automated Specification) es uno de los diagramas más importantes durante la educación de requisitos de software, es decir, de la primera fase del ciclo de vida del software porque permite expresar a los interesados (usuarios) la importancia del software futuro. En los trabajos que utilizan este diagrama no se logra identificar la trazabilidad que debe existir entre el lenguaje natural y los objetivos identificados para evitar inexactitudes entre éstos. En este artículo, se presenta un método para la extracción y clasificación de objetivos bajo el enfoque KAOS a partir del procesamiento de requisitos textuales en lenguaje natural del idioma español. En este método, se efectúa la descomposición de las oraciones que se encuentran presente en un texto y se realiza el análisis morfológico y sintáctico de cada oración, tomando como referencia las clasificaciones en español de los verbos para obtener objetivos. Además, se definen estructuras morfosintácticas que permiten tipificar y clasificar objetivos de acuerdo con los siguientes tipos: (i) Mantenimiento; (ii) Logro; (iii) Terminación; y (iv) Evasión. El propósito de esta clasificación está orientado a listar los objetivos y representarlos según el modelo de objetivos KAOS. Este proceso sirve como punto de partida para la identificación de los demás elementos que componen el diagrama de objetivos de KAOS y la elaboración semiautomática de este.*

*Palabras clave: Metodología KAOS, diagrama de KAOS, identificación de objetivos, ingeniería de requisitos, procesamiento del lenguaje natural.*

<sup>1</sup> Departamento de la Decisión y Computación, Facultad de Minas. Universidad Nacional de Colombia. Medellín, Colombia.  
E-mail: lalezcan@unal.edu.co; jaguzman@unal.edu.co; seagomezar@unal.edu.co

## INTRODUCTION

Lamsweerde and Letier [1] show that KAOS is a methodology used in Requirement Analysis in order to carry out software requirement elicitation. [2] indicates that the KAOS methodology allows analysts to identify requirements in any information system. The main advantage of this methodology is the ability to align the organization's requirements, goals and expectations.

The KAOS methodology is based on a goal-oriented approach which presents various levels of expressiveness and reasoning: (i) semi-formal level for modeling and structuring goals; (ii) qualitative level for choosing between alternatives and (iii) formal level (if necessary) for more precise reasoning concerning the different components associated with the requirements [3].

Each component (goal, expectation, requirement) in the KAOS modeling language has a two-level structure: the external layer (semantic and graphic) where the concept is stated along with its attributes and its relations with other concepts, and the internal layer for formally defining the concept [4]. Currently it is the analyst that manually and subjectively creates the goal diagram using information provided by the stakeholder, without receiving any help in conceptualizing said diagram [5]. Taking this above into account, it is necessary to automate this process in order to obtain a KAOS goal diagram from natural language.

The paper is set out as follows: the theoretical framework for contextualizing the KAOS model is set out in next section; then the components used for processing natural language and structuring the methodology, such as morphosyntactic structures and verb classification, are detailed; next, the methods for obtaining goals and for the experiments carried out are presented, while in last section the conclusions and future studies derived from this article are laid out.

## THEORETICAL FRAMEWORK

The KAOS methodology states that goals may be classified as Maintain, Achieve, Avoid and Cease goals. Each one has a specific notation within the KAOS goal model [1]. See Figure 1.

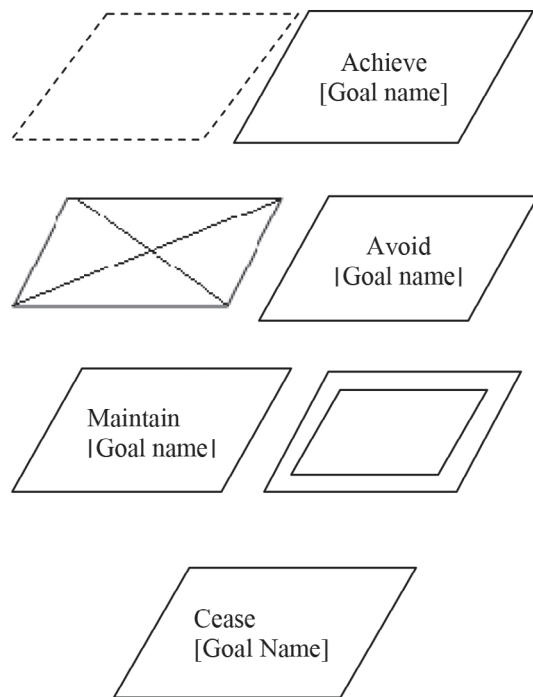


Figure 1. Representation of goal types under the KAOS model. Source: Taken from [1].

The KAOS methodology defines a model for representing system requirements [6]. This representation is made up of the following models: (i) goals. Characterized mainly by the definition of the organization's goals, requirements and expectations as well as their hierarchy; (ii) objects. Characterized by representing the organization's low-level components, such as the relation between entities from the UML classes model; (iii) agents. Characterized by the elements responsible for carrying out an action within the organization and their representation; (iv) operation. Characterized by the specification of the goals at the operationalizable level, i.e. the reduction of complex utterances (such as goals) to simple operations.

## NATURAL LANGUAGE PROCESSING (NLP)

NLP is one of the branches of Artificial Intelligence (AI) which allows information from a domain to be obtained from a discourse. This information may be formalized by an inference mechanism employed on a particular text or discourse [7]. According to the grammar of cases from [8], expected behavior in language can be modeled from a concrete set of

cases. It is because of this that the method proposed here will make use of this approach. In other words a set of cases using different forms which attempts to completely extend them to the language. Generally a sentence in Spanish which describes an object follows a basic structure that can be used to identify and classify an object by making use of a syntactic analysis. The components which make part of the identification process for this structure are listed below:

Periphrasis: these are syntactic constructions of two or more verbs which function as a predicate. They serve to express the characteristics of the verbal action that cannot be indicated through the use of simple or composite forms [9]. Their macro structure can be seen in Table 1.

Table 1. Macro structure of the periphrases.  
Source: Taken from [9].

Auxiliary verb	Joining word	Auxiliated verb
Provides the grammatical content. Appears in personal form.	They may be conjunctions or prepositions, or there may not be a joining word.	Provides semantic content. Appears in impersonal form (infinitive, gerund, participle).

Likewise the periphrases can be grouped in two large groups: aspectual periphrases and modal periphrases. The latter group is of real interest as they denote actions, tasks and desires. See Table 2.

Table 2. Micro structure of the modal periphrases.  
Source: Taken from [9].

Modal Periphrases		
Obligation	Tener que/ Deber/ Haber de / Haber que	+ Infinitive
Probability or suspicion	Venir a / Deber de	+ Infinitive
Possibility	Poder	+ Infinitive

Noun Phrase: this is a group of words which articulate around the noun. In this syntagma it is the nucleus [10].

Nominalized Verb: this is the noun which is derived from the verb (eg. la *realización* (realization) which comes from *realizar* (to realize).

Complement: in this context it is the group of words which support the sentence or make up the sentence in question, with the aim of lending meaning or significance.

### Key components in defining goals

The method proposed here utilizes verb classification for goals characterized in [11]. Moreover it defines 4 morphosyntactic forms in which a goal can be described using a sentence in Spanish. These forms are shown below.

### Identifying a goal's morphosyntactic forms in Spanish

The following proposed grammatical rules allow a goal to be characterized in Spanish independently of the type to which it belongs (according to the KAOS classification).

**F1:** Noun Phrase + Modal Periphrasis + Nominalized Verb + Complement.

**F2:** *Que* ('That', *Optional*) + Noun Phrase + Verb + Complement.

**F3:** Verb in the Infinitive + *que* + Noun Phrase + Verb + Complement.

**F4:** Noun Phrase + Modal Periphrasis -Complement.

In Table 3 one can see an example of these rules.

### Classification of verbs according to the KAOS goal types

One of the main components of a sentence in Spanish is the verb. As a result and based on context, the definition and characterization of a verb can completely determine sentence meaning. Given the above, characterizing a verb becomes crucial in being able to adequately classify a sentence. Verbs in Spanish can be grouped into four categories according to their lexical features. In this context their lexical characteristics (Telicity, Dynamism, Durability) may be classified into four groups [11]. See Table 4.

Table 3. Example of the previously described forms.  
Source: authors' own work.

Spanish Natural Language	Form	Graphical Representation In KAOS
El sistema debe permitir el registro de los clientes. (The system should allow for client registration)	F1	Permitir el registro de los clientes (Allow the clients registration)
Que la secretaria obtenga los teléfonos. (Let the secretary obtain the phone numbers)	F2	Obtener los teléfonos (Get the phone numbers)
Garantizar que el director guarde los clientes. (Make sure that the manager stores customer information)	F3	Guardar los clientes (Save the clients)
El sistema debe validar los datos. (The system should validate data)	F4	Validar los datos (Validate data)

[12] indicates that a high-level goal for an IT solution should be classified into one of the following goal types: Achieve, Maintain, Cease and Avoid. These in turn have a characteristic verb type associated with their definition in one or some of the types described above. See Table 5.

Table 4. Characterization of Spanish verbs according to their lexical characteristics.

Source: Taken from [11]

Verb Type	Telicity	Dynamism	Durability	Example
Estado (State)	–	–	+	Tener, (keep) Poseer (have)
Actividad (Activity)	–	+	+	Marchar, (march) Hablar (talk)
Realización (Realization)	+	+	+	Destruir, (destroy) Construir (build)
Logro (Achievement)	+	+	–	Alcanzar (achieve), Notar (note)

Table 5. Classification of goal verbs in Spanish. (Adapted from [11]).

Verbs for different goal types			
Achieve	Maintain	Cease	Avoid
Formar (form)	Administrar (Administer)	Parar (Stop)	Evitar (Avoid)
Mejorar (Improve)	Garantizar (Guarantee)	Finalizar (Finish)	Obstaculizar (Block)
Aumentar (Increase)	Guardar (Save)	Suspender (Stop)	Prevenir (Previce)
Fomentar (Promote)	Ofrecer (Offer)	Detener (Stop)	Esquivar (Avoid)
Registrar (Register)	Prolongar (Prolong)	Concluir (end)	Eludir (Avoid)
Elaborar (Develop)	Avalar (Endorse)	Interrumpir (Interrupt)	Sortear
Formular (Formulate)	Gestionar (Manage)	Cesar (Cease)	
Promover (Promote)	Obtener (Obtain)	Terminar (End)	
Hacer (Make)	Preservar (Keep)		
Meter (Insert)	Perdurar (Keep)		
Obrar (Make)	Conservar (Keep)		
Preparar (Prepare)	Mantener (Keep)		
Reducir (Reduce)	Reconocer (know)		
Decrecer (Reduce)			
Incrementar (Increase)			

### Proposed method for obtaining and classifying goals using natural language

The classification of the verbs used in Table 5 will be used in this proposal to classify the goals. This means that once the grammatical rule is identified

(see identifying a goal's morphosyntactic forms in Spanish) which is applicable to a sentence, the verb classification will be identified based on verbs shown in Table 5.

The stakeholder is the party that explains the domain problem to the software analyst. This description is used as an input to begin the goal identification and classification process based on the KAOS methodology, and the description can be obtained from documents written in natural language by the stakeholder or analyst. The input under the framework for this study will be a document written in natural language in Spanish which contains the textual requirements of a system. The morphosyntactic structure of the text that one hopes to process is restricted to utterances in the active voice, as they clearly indicate the subject of the action. The text should also be well-written with regards to punctuation and orthography, and all phrases should explicitly contain the subject of the phrase. This is a simplification which is performed with the aim of avoiding working with the anaphora resolution, an additional problem which falls outside the scope of the proposed method.

The method consists of 8 consecutive steps. A computational tool was created in JAVA and PHP for this, which allows results to be obtained automatically after inserting the natural-language document with the syntactic tagging performed in Freeling 3.0 [13]. These steps are listed below. Figure 2 shows the activities diagram laid out in columns for the proposed model (Square means activity/process and cylinder represent knowledge source/rules/data). This also involves the NL2KAOS computational tool developed for this proposal.

Each one of the steps that must be followed is related below:

Step 1: introduce the text of the description (which will serve as an initial process input) into the NL2KAOS computational tool developed within the framework of this proposal. Step 2: identify and store the sentences contained in the text **in a list** based on its punctuation (commas or periods). Step 3: eliminate empty words (articles, determiners and conjunctions) from the original text, as is the case. Step 4: the dependency tree provided by Freeling 3.0 [13] is attained for each sentence obtained in

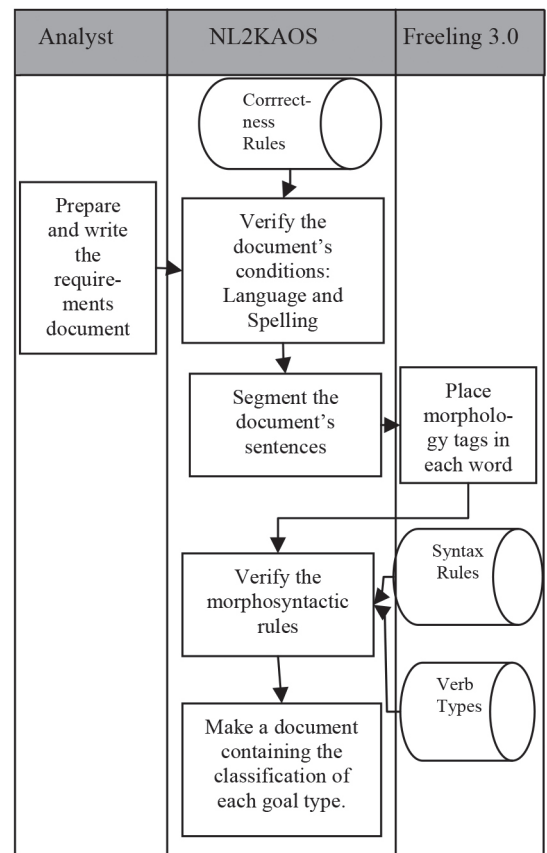


Figure 2. Proposed activities diagram for obtaining the KAOS goal types. Source: authors' own work.

step 2, and this classification is compared with each one of the forms established in the previous section. Step 5: each sentence which complies with step 4 will be stored in the `{possible_objetivos1}` **list**, and those which do not will be stored in the `{no_objetivos}` **list**. Step 6: for each sentence stored in `{no_objetivos}`, goal verbs described in section 3.3 or their synonyms are searched for using Multilingual Central Repository [14]. Sentences where this type of relation between verbs is found are stored in `{possible_objetivos2}`. Step 7: search for target verbs or their synonyms in the sentences contained in `{possible_objetivos1}`. Once identified they will be stored in `{objetivos_definitivos}`. Once these 7 steps have been followed a set of phrases is obtained and stored in `{objetivos_definitivos}`. One then proceeds to the next step. Step 8: the verbs are identified for each utterance present in `{objetivos_definitivos}` and the goal verb type found in the utterance is searched for and stored



in {objetivos\_evasion}, {objetivos\_terminacion}, {objetivos\_logro} and {objetivos\_mantenimiento}, as is the case. Thus four lists will be obtained, each with one goal type.

## EXPERIMENTS

The proposed method was assessed with three different classic utterances from Software Engineering, to which the whole of the previously described process was applied. For validation purposes, only the experiment carried out involving “the elevator case study” proposed in [2] was presented for this paper, due the syntactic rules was made just for Spanish language we present the text in Spanish, please review [2] to view the original text in English.

### Elevator case study Spanish version, adapted from [2]

*“La compañía encargada del elevador está en la capacidad de proveer una forma de escape. La compañía debe elaborar una interfaz basada en botones. La compañía debe garantizar la existencia de un botón de emergencia. La compañía debe elaborar la estructura del ascensor. La compañía debe garantizar la energía de emergencia. La compañía debe garantizar el funcionamiento del software.*

*El controlador del elevador debe garantizar que las puertas no se abran mientras esté en movimiento. El controlador debe detener el elevador, si hay una falla de energía. El controlador debe encender la luz de emergencia cuando sea necesario. El controlador puede abrir las puertas, cuando esté en el nivel indicado. El controlador debe informar a los pasajeros del estado de su petición. El controlador debe reportar a los pasajeros las condiciones de sobrepeso”.*

### English version

*“The company in charge of the elevator is able to provide some means of escaping. The company should create a button-based interface. The company should guarantee the existence of an emergency button. The company should make the structure of the elevator. The company should guarantee the availability of emergency power. The company should guarantee that the elevator’s software works.*

*The controller should guarantee that the doors will not open while in motion. The controller must*

*stop the elevator if there is a power shortage. The controller should turn on an emergency light when necessary. The controller can open the doors when the elevator is in the appropriate floor. The controller should inform passengers of the status of their request. The controller should inform passengers of any overweight conditions.”*

### Result Analysis

For the manual process carried out by the analyst using the proposed rules, 12 goals were identified. Of these, 7 were classified in the following way: 4 maintain, 2 achieve and 1 avoid. Using the automatic proposal presented in this paper one was able to obtain the same quantity of goals and agents with relation to those identified manually. This leads us to conclude that, in this proposal, one was able to comply with: (i) the completeness characteristics with relation to the number of components that were identified; (ii) the consistency (suitable for this specification), i.e. the traceability that should exist between the natural language and the KAOS goals. These were identified in less time given that manual process took 12 minutes, whereas the automatic process took 12 seconds. The results can be seen in Figures 3 and 4.

The tools used for the automatic test were: (i) Hardware (Intel CORE i3 - 2.4 GHz processor with 3 GB de RAM); and (ii) Software (Ubuntu Desktop 12.1, Freeling 3.0, MultiWordNet and NL2KAOS (developed in this proposal)).

## CONCLUSIONS

Four morphosyntactic forms were conceptualized for this study for identifying goals from common identifiable components in natural language. A goal classification, based on verb types found in Spanish, was also defined in the description of a system.

A method for obtaining goals for the KAOS model by processing a natural language discourse was created. This method takes as its basis the comparison of morphological structures found in Spanish phrases taken as goal forms.

The method was also automated using the NL2KAOS computational tool created in JAVA and PHP. The tool was used to process natural language from Freeling 3.0 and the Multilingual Central Repository [14] lexicon in order to tag and find synonymy between

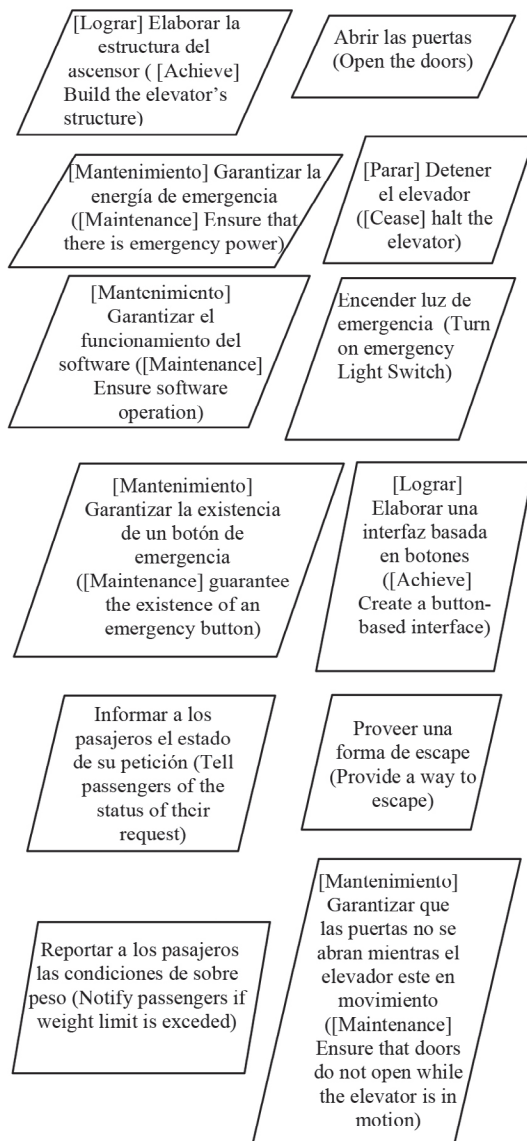


Figure 3. Goals identified in the case of the elevator.  
Source: authors' own work.

words. These tools allowed the methodology to be carried out so as to eventually assess the methods using three classic utterances from Software Engineering, and present the results.

This paper has given rise to new topics for study which may give continuity to this study. Some of these are listed below: (i) define more morphosyntactic forms of representation for goals; (ii) characterize and include more verbs which could be synonyms for any of the four verb types presented in this study;

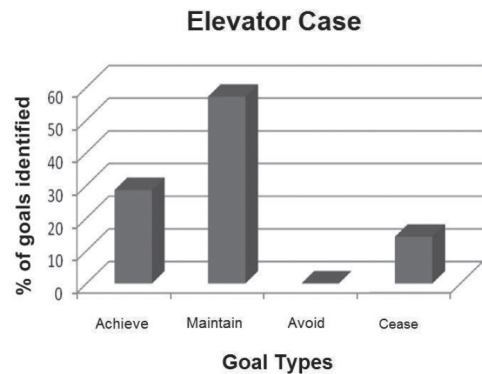


Figure 4. Percentage of goals identified in the case of the elevator.

(iii) define additional rules which could generate variations in the goals identified through this study; and (iv) resolve the anaphora resolution.

## ACKNOWLEDGEMENTS

This paper is part of a research Project entitled “A Terminological Processing Model for Obtaining Software Requirements Based on the KAOS Goal Diagram”. The project's code is 202010011022, and it is sponsored by the Research Directorate of Universidad Nacional de Colombia - Sede Medellín (DIME) through the “DIME 2012 Research Project Financing” program.

## REFERENCES

- [1] A. Lamsweerde and E. Letier. “Object Orientation to Goal Orientation: A Paradigm Shift for Requirements Engineering”. *Radical Innovations of Software and Systems Engineering in the Future*, pp. 153-166. Springer. Venecia, Italia. 2004.
- [2] Respect IT. “A KAOS Tutorial”. Objectiver 2007. Date of visit: July 7, 2013. URL: [http://www.objectiver.com/file\\_admin/download/documents/KaosTutorial.pdf](http://www.objectiver.com/file_admin/download/documents/KaosTutorial.pdf)
- [3] F. Almisned and J. Keppen. “Requirements Analysis: Evaluating KAOS Models”. *2nd International Workshop for Requirements Analysis*, pp. 869-874. Londres, Reino Unido. 2010.
- [4] A. Lapouchnian. “Goal-Oriented Requirements Engineering: An Overview of the Current Research”. University of Toronto. Toronto, Canadá. 2005.



- [5] C. Zapata, S. Villegas and F. Arango. "Reglas de consistencia entre modelos de requisitos de UN-Método". University Eafit Journal. Vol. 42, Issue 141, pp. 40-59. 2006.
- [6] A. Lamsweerde. "Requirements Engineering. From System Goals to UML Models to Software Specifications", pp. 278-290. Great Britain, Reino Unido. 2009.
- [7] A. Cvitas. "Relation Extraction from Text Documents. MIPRO". Proc. 34th International Convention, pp. 1565-1570. Opatija, Croacia. IEEE. 2011.
- [8] C. Fillmore. "Universals in Linguistics Theory". Vol. 1. E.B. Harms, Ed. Holt, Rinehart and Winston Publishing Company. 1968.
- [9] F. Genta. "Perífrasis Verbales en español: Focalización aspectual, restricción temporal y rendimiento discursivo". PhD. Thesis, pp. 165-250. Universidad de Granada. España. 2008.
- [10] D. Violeta and B. Ignacio. "Gramática descriptiva de la lengua española". Espasa C. España. 1999.
- [11] C. Zapata and L. Lezcano. "Characterization of Goal Diagram Verbs". Journal Dyna. Vol. 76, Issue 158, pp. 219-228. 2008.
- [12] A. Lamsweerde, A. Dardenne, B. Delcourt and F. Dubisy. "The KAOS Project: Knowledge Acquisition in Automated Specification of Software". Proceedings AAAI Spring Symposium Series, Stanford University, American Association for Artificial Intelligence, pp. 59-63. 1991.
- [13] FreeLing 3.0 "An open source suite of language analyzers". Date of visit: July 4, 2013. URL: <http://adimen.si.ehu.es/web/MCR/>
- [14] A. Gonzalez, E. Laparra and G. Rigau. "Multilingual Central Repository version 3.0". 8th international conference on Language Resources and Evaluation (LREC'12). Istambul, Turkey. 2012. Date of visit: July 5, 2013. URL: <http://adimen.si.ehu.es/web/MCR/>