

# Aprendizaje de Reglas Encadenadas para la Creación de Grafos Conceptuales

Sonia Ordoñez Salinas

**Resumen**—El documento presenta una forma de aprendizaje sobre reglas encadenadas para la generación de nuevas reglas que al aplicarlas deberán permitir la construcción de Grafos Conceptuales. La propuesta se basa en la inclusión de reglas encadenadas y de un método supervisado. Las reglas son definidas sobre la base de tres elementos: a)La marcación o rol que ocupa la palabra dentro de la oración, b)El estándar de Grafos Conceptuales y c) La definición de un Objeto que funciona como una caja de negra de Grafos. Las pruebas se realizaron sobre algunos de los textos correspondientes a los títulos y comentarios que hacen parte de la colección de imágenes médicas del ImageClefmed del 2008. Para la realización de las marcas se utilizó el metatesauro UMLS y la herramienta MMTx y para los procesos de clasificación se uso el Weka. Como resultado se estiman nuevas reglas.

**Palabras clave**—Aprendizaje de reglas encadenadas, UMLS, grafos conceptuales, anotación de imagen médica.

## Learning of Chained Rules for Construction of Conceptual Graphs

**Abstract**—El documento presenta una forma de aprendizaje sobre reglas encadenadas para la generación de nuevas reglas que al aplicarlas deberán permitir la construcción de Grafos Conceptuales. La propuesta se basa en la inclusión de reglas encadenadas y de un método supervisado. Las reglas son definidas sobre la base de tres elementos: a)La marcación o rol que ocupa la palabra dentro de la oración, b)El estándar de Grafos Conceptuales y c) La definición de un Objeto que funciona como una caja de negra de Grafos. Las pruebas se realizaron sobre algunos de los textos correspondientes a los títulos y comentarios que hacen parte de la colección de imágenes médicas del ImageClefmed del 2008. Para la realización de las marcas se utilizó el metatesauro UMLS y la herramienta MMTx y para los procesos de clasificación se uso el Weka. Como resultado se estiman nuevas reglas.

**Index terms**—Aprendizaje de reglas encadenadas, UMLS, grafos conceptuales, anotación de imagen médica.

## I. INTRODUCTION

La necesidad de una comunicación efectiva entre el hombre y la máquina, de tal forma que una persona pueda expresarse en su Lenguaje Natural y la máquina pueda responderle en el mismo lenguaje es una necesidad cada vez más requerida. Ningún dominio se escapa a esta necesidad y

Manuscrito recibido el 15 de abril del 2010. Manuscrito aceptado para su publicación el 14 de junio del 2010.

La autora trabaja en Universidad Distrital F.J.C, Colombia y estudia en Universidad Nacional de Colombia (e-mail: sordonez@udistrital.edu.co).

en particular el dominio médico. Las personas involucradas con la asistencia médica generalmente realizan sus anotaciones en lenguaje natural y por más que exista software especializado y muy estandarizado, a la hora de expresar los síntomas, el diagnóstico de un paciente o la anotación sobre una imagen médica, el personal médico simplemente lo digita en su lenguaje natural.

A pesar de que existen herramientas especializadas como las ontologías y los tesauros que ayudan a comprender algunos de los términos digitados como nombres de enfermedades y algunas relaciones entre conceptos, es claro que poder procesar el lenguaje natural por el computador es una tarea que no es fácil. Si bien para el ser humano, ante una frase, inmediatamente no solo la relaciona con todo su conocimiento previamente adquirido sino que es capaz de filtrar todo ese conocimiento, para ubicar la frase en el entorno o dominio adecuado y dejar exactamente lo que necesita para saber de qué se trata. Los procesos utilizados con la máquina o el computador no pueden hacer esto tan inmediato, ni con tan buen resultado, ante una frase solo pueden ubicar una cadena de símbolos que no significan nada, por lo que se debe recurrir a ontologías o tesauros y consultar que pueden significar estos símbolos. Por otro lado, para poder individualizar un símbolo o un conjunto de estos y sus relaciones se requiere que sean representados dentro de estructuras más estándares y adecuadas para el computador. Adecuadas en el sentido que se pueda gestionar uno o varios de estos símbolos, es decir, almacenarlos, consultarlos y eliminarlos.

Existe una gran variedad de estructuras que se utilizan para realizar la tarea de procesar el texto a través de un computador, unas con mayores ventajas que otras. Sin embargo todas cuentan con la dificultad de requerir pasar los símbolos y/o conocimiento (cuando el símbolo ya tiene algún significado) a dicha estructura. Generalmente se utiliza la ayuda de un experto para representar el texto expresado en lenguaje natural en una de estas estructuras.

Dentro de estas estructuras se encuentran Los Grafos Conceptuales que a diferencia de las demás estructuras, permiten incluir detalles de la semántica propia del texto escrito en Lenguaje Natural y cuentan con algunas características del lenguaje matemático.

Este trabajo se ocupa en forma particular, de cómo automáticamente generar nuevas reglas a partir de un conjunto suministrado de reglas encadenadas. La idea de contar con un conjunto de reglas que permitan la consecución de manera

automática de Grafos Conceptuales a partir de un texto expresado en lenguaje Natural.

La temática de cómo crear grafos a partir de un texto, se ha tratado de diferentes enfoques desde aquellos que implican la consecución de cada uno de los elementos de manera secuencial hasta aquellas que involucran procesos estadísticos. A pesar de las diferentes propuestas en torno al fin planteado y de acuerdo al estado del arte, se puede observar que cualquier esfuerzo encaminado a dar nuevas alternativas aún es válido.

Este documento, presenta la propuesta de encontrar reglas a partir de reglas encadenadas por medio un método supervisado. Las reglas encadenadas al ser aplicadas sobre un texto expresado en lenguaje natural, dan como resultado la representación de dicho texto en una estructura de Grafo Conceptual. Las reglas encadenadas son encontradas con base en tres elementos: a) el rol que ocupa una palabra dentro la frase b) las estructuras básicas de los Grafos Conceptuales y c) la definición de un *Objeto* que funciona como una caja negra de grafos pero que da la posibilidad de inferir las reglas inmersas dentro de esta caja negra. Los textos son extractados de las anotaciones que hacen parte de la colección de imágenes médicas del ImageClefMed del 2008.

Este documento se organiza de la siguiente forma: en la sección II se definen los conceptos básicos incluyendo los lineamientos generales sobre el tema tratado, permitiendo ubicar al lector. En la sección III se incluye una revisión del estado del arte. En la sección IV se presenta el método utilizado; en V se muestra el desarrollo experimental. En VI se presentan las conclusiones y trabajo futuro.

## II. CONCEPTOS BÁSICOS

El dominio de la medicina cuenta con herramientas muy sofisticadas para ayudar a los sistemas computacionales con la terminológica propia del dominio. Dentro de estas ayudas especializadas se encuentra el Metatesauro “*Unified Medical Language System (UMLS) of the National Mibrary of Medicine of the National Institute of Health of United States*” [20], desarrollado en el 2003. Dicha herramienta consta de una base de datos multilingüe que relaciona más de un millón de conceptos biomédicos, de la salud y léxicos, incluidos en más de 100 fuentes de información. Cuenta además, con una serie de programas especializados para manejar dicha información. Dentro de la herramientas que hace parte de la colección está el MetaMap Transfer (MMtx) [13] que permite entre otras funciones, relacionar términos del metatesauro UMLS y hacer marcación del rol que ocupa una palabra dentro de la Oración (*Part-of-speech tag*). Esta herramienta a diferencia de otras de su categoría es que está hecha específicamente para que funcione con el metatesauro UMLS y que por ende, reconoce las palabras propias de las ciencias biomédicas y de la salud. Es así, que no solo reconoce una palabra o conjunto de palabras sino que permite etiquetar estas palabras.

A pesar de la existencia de herramientas como el UMLS aún se hace necesario la utilización de otras técnicas que permitan suplir las necesidades relacionadas con el análisis del lenguaje y que son propias de la actividad del personal médico y hospitalario.

El Procesamiento del Lenguaje Natural (PLN) es una disciplina encargada de la gestión del Lenguaje Natural (LN). Entendida como gestión la extracción de información a partir de un texto expresado en lenguaje natural y su procesamiento [8]. El proceso incluye desde almacenar e indexar la información hasta la búsqueda y la consulta de un requerimiento en particular. Dicho procesamiento es posible gracias a una serie de artificios o estructuras más acordes para el ámbito computacional que un texto escrito en lenguaje natural. Dentro de estas estructuras se mencionan algunas a continuación.

### A. Estructuras Matemáticas

Las estructuras matemáticas son las más utilizadas para el procesamiento del lenguaje natural. Dentro de estas se pueden referenciar las estructuras vectoriales que pueden ser utilizadas con números o palabras. Trabajar con estas estructuras es muy fácil, sin embargo, tienen la desventaja que al representar un texto escrito en LN en dichas estructuras permiten que se pierdan muchos de los detalles propios de la semántica inmersa dentro del lenguaje natural. Se puede, por ejemplo relacionar el trabajo presentado en [6], que utiliza los vectores y el presentado en [18], que utiliza una estructura más compleja como son los grafos. En esta categoría se pueden incluir la clásica estructura para modelamiento de bases de datos y conocida como Modelo Entidad Relación y el modelo de clases utilizado para el diseño de aplicaciones de software.

### B. Lenguajes Estructurados

Utilizados especialmente para describir los pasos y procedimientos que debe incluir un programa de software o los conceptos y funcionalidades propias de una ontología.

Los lenguajes clasificados bajo esta categoría se definen como metalenguajes orientados al reconocimiento y descripción de objetos y clases. Dentro de esta categoría se encuentran la mayoría de lenguajes de programación orientados a objetos, donde la unidad primaria de organización es un marco que tiene características como el nombre y los atributos. Un atributo igualmente tiene su nombre y puede contener valores.

### C. Estructuras Gramaticales

Las estructuras gramaticales permiten incluir la información gramatical producto de los analizadores gramaticales y generalmente es expresada en forma de árboles y reglas gramaticales. Dentro de estas estructuras estas las reglas utilizadas para expresar las gramáticas libres de contexto.

### D. Estructuras Conceptuales

Las estructuras propias de esta categoría, a diferencia de las anteriores permiten incluir una mayor riqueza semántica.

Dentro de estas estructuras se pueden catalogar, las redes semánticas [13], los lenguajes para descripción e intercambio de conocimiento como el *Knowledge Interchange Format* (KIF), el *Resource Description Framework* (RDF), el *web ontology language* (OWL) y los Grafos Conceptuales [12]. Estos últimos fueron diseñados para que su representación incluyera a demás detalles propios de la semántica, algunas de las bondades propias de las matemáticas.

### III. ESTADO DEL ARTE

En esta sección se mencionan las características principales de los Grafos Conceptuales y se relacionan algunos de los trabajos sobre cómo de manera automática representar un texto escrito en el Lenguaje Natural en un Grafo Conceptual. Dentro de los trabajos se presentan algunos que se enfocan a la parte procedural o metodológica, a procesos puramente secuenciales y aquellos que de alguna forma utilizan aprendizaje maquinal.

#### A. Grafos Conceptuales

Los grafos conceptuales son definidos por Sowa [19] con base en la lógica de primer orden y los grafos existenciales de Pierce [6]. El estándar incluye dos tipos de nodos (conceptos y relaciones), aristas y expresiones de tipo lógico. Por sus características los Grafos Conceptuales no solo, permiten incluir detalles propios del lenguaje natural, sino que soportan operaciones básicas como la unión, la intersección y la inferencia. El grafo correspondiente a “*La radiografía enseña un corazón crecido*” se puede visualizar en la Figura 1.

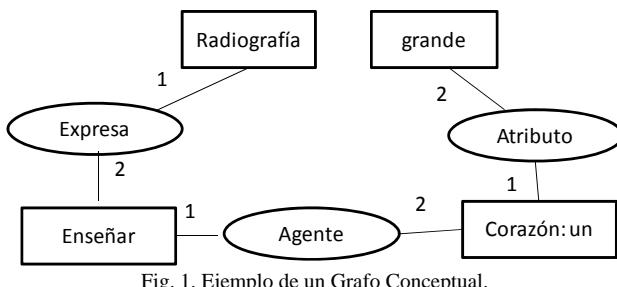


Fig. 1. Ejemplo de un Grafo Conceptual.

#### B. Representación del Lenguaje Natural en Grafos Conceptuales

En cuanto a la forma cómo pasar un texto expresado en lenguaje natural a una estructura de Grafo Conceptual se han propuesto diferentes métodos, entre los que siguen pasos secuenciales y procedimentales para la consecución de cada uno de los elementos propios de los Grafos Conceptuales y aquellos que incluyen procesos de aprendizaje automático.

1) *Métodos Secuenciales y Procedimentales*. Dentro de los trabajos que se enfocan a la parte procedural o metodológica se pueden citar el presentado [17] que es puramente procedural y el [15] que presenta la creación de los Grafos Conceptuales desde el punto de vista de modelamiento remedando los lineamientos que se siguen al crear un modelo de clases.

Como trabajo que sigue una serie de pasos de manera secuencial se encuentra el [5] que presenta un sistema que

convierte un texto en español a Grafos Conceptuales, utilizando un analizador morfológico (véase, por ejemplo, [8]), un analizador sintáctico e identifica cada uno de los elementos de los grafos.

2) *Métodos que Involucran Tareas de Aprendizaje Automático*. En el trabajo presentado en [1] se obtienen los Grafos Conceptuales a partir de textos en francés basado en algunas estructuras y el análisis sobre los verbos.

Sobre el dominio específico de la medicina se encuentran trabajos como el presentado en [16], donde se busca construir una estructura de reportes basada en Grafos Conceptuales. Con tal fin, se analizan las frases a partir de reglas de proximidad que tienen en cuenta la información semántica y sintáctica obteniendo un texto anotado. Con base a dicho análisis se consiguen los componentes que permiten construir los Grafos Conceptuales.

En el propuesto en [19], se construye un sistema que se enmarca en las radiologías y se representa el contenido de las anotaciones de las radiologías en Grafos Conceptuales.

El trabajo presentado en [4], con el objeto de lograr un mejor desempeño a la hora de realizar consultas y recuperación, presenta algunos aspectos de la transcripción de registros de asistencia médica a una estructura de Grafos Conceptuales. El trabajo se centra en el modelamiento de una base de conocimiento utilizando Grafos Conceptuales, muy similar a como se modela bajo el paradigma *entidad-relación*.

Utilizando reglas sintácticas y semánticas se puede citar el trabajo presentando en [10] y el planteado en [2] que utilizan un analizador gramatical para construir árboles gramaticales y a partir de una serie de reglas transforman dichos árboles en grafos conceptuales. Una vez transformados se optimizan los grafos al desambiguarlos estructural y semánticamente.

En la investigación presentada en [23], construyen los Grafos Conceptuales con base en gramáticas de enlace y máquinas de aprendizaje como un problema de clasificación que puede ser entrenado para diferentes dominios.

### IV. MÉTODO UTILIZADO

Con el fin de conseguir de manera automática los Grafos Conceptuales, se propone el aprendizaje maquinal a través de un conjunto de reglas encadenadas. Cada clase de regla aplicada sobre un texto fue etiquetada obteniendo al final una cadena de reglas encadenadas. La idea es que, al aplicar un conjunto de reglas de forma encadenada y sobre un texto particular se obtiene el Grafo Conceptual correspondiente a dicho texto

#### A. Colección Experimental

La colección de entrenamiento y prueba se conforman a partir de un conjunto de texto en lenguaje natural y relacionado con anotaciones de imágenes médicas de la colección de ImageClefMed del 2008. Los textos fueron extractados tanto del título como de la anotación de la imagen.

De algunos de los textos de esta colección y utilizando el software MetaMap Transfer (MMTx) se etiquetaron los textos con las marcas correspondientes a la función que juega cada palabra dentro de la oración (ver Tabla I).

**TABLA I**  
EJEMPLO DE LAS ETIQUETAS UTILIZADAS POR MMTx.

| Etiqueta de MMTx | Significado  |
|------------------|--------------|
| noun             | Sustantivo   |
| adj              | Adjetivo     |
| prep             | Preposición  |
| det              | Determinante |
| aux              | auxiliar     |
| ver              | Verbo        |

A esta colección se le realizó un preprocesamiento previo con el fin de facilitar las tareas siguientes. Dentro de este preprocesamiento se resolvió la correferenciación y se asumió:

*Sustantivo: un sustantivo equivalente a un conjunto de* sustantivos consecutivos. Esto es, que en el caso de encontrar más de dos sustantivos consecutivos se asumen como si fuera un único sustantivo. Lo anterior permitió definir las entidades nombradas, como “*Seymour H. Levit, MD Radiology Society of North America*” (en este caso la herramienta MMTx detectó “*Society of North America*”).

*Concepto Básico (CB):* Un sustantivo con características definidas por el estándar de Grafo Conceptual (ver Tabla II).

En el Lenguaje Natural, una idea se puede expresar de diferente forma y para el raciocinio del ser humano dicha idea significa lo mismo. Para el ámbito computacional, esto no es tan simple. Existe la técnica computacional llamada Parafraseo que consiste en cambiar una frase por otra sin alterar su significado. Este cambio se puede dar en dos sentidos, el primero se cambia algunos términos por sinónimos y en el segundo se cambia totalmente la forma de la oración. En esta última forma, se pueden cambiar todas o algunas palabras y el orden de las mismas, por ejemplo pasar una oración de forma activa a su forma pasiva. Por lo que encontrar si dos oraciones son similares se debe aplicar la técnica de parafraseo a una de ellas con el fin de ver si de alguna forma se llega a la otra.

Los grafos conceptuales como una estructura estándar busca representar algunas ideas de manera estándar o que una idea siempre se represente de igual forma. Es así, con el caso de la presentación de un sustantivo, incluido la categoría del sustantivo y su individualización. Por ejemplo la frase “Paul is a child”, puede escribirse de varias formas entre ellas: “Paul who is a child”, “the child is Paul”, “the child name is Paul”. Para los Grafos Conceptuales, la idea se representa a través del nodo concepto [child:Paul]. Para esta estandarización un nodo concepto debe estar formado por mínimo la categoría. Esto es, que si no se conoce el nombre del niño el grafo sería [child]. Pero si solamente se conoce que es una persona, animal o cosa se llama Paul, el concepto no se puede representar como [Paul], ya que es completamente ambiguo, por lo que se debe incluir la categoría y esto requiere de un proceso de búsqueda el cual no es objeto del presente trabajo.

Para resolver la categoría se definió la categoría “nn” permitiendo representar un concepto de la forma [nn:Paul].

**TABLA II**  
EJEMPLO DE GC DE TIPO CONCEPTO BÁSICO.

| Paul is a child      | Notación GC  |
|----------------------|--------------|
| That child           | [niño:Paul]  |
| Some child           | [niño:#That] |
| Every child          | [niño:{*}]   |
| There exists a child | [niño:□]     |
| One cat              | [cat:@'one]  |

**Objeto (Obj):** Un elemento que puede ser concepto o el resultado de aplicar una regla y que no es otra cosa que un uno o varios Grafos Conceptuales compuestos por nodos concepto, nodos relaciones y aristas. Este Objeto se define para que funcione como una caja negra que incluye un grafo o n-grafos y sobre la cual se puede aplicar una función que da como resultado otra caja negra.

**Regla:** Las Reglas se definen bajo los estándares de los Grafos Conceptuales. Al aplicarlas equivalen a las relaciones entre los nodos. De acuerdo al estándar de los grafos conceptuales y siguiendo con la idea del parafraseo, se busca que estandariza hasta donde sea posible la representación de varias ideas. Es el caso de los verbos, donde algunos emiten la idea de una percepción, entonces, todos las ideas que incluyan una percepción siempre se deberá representar de igual forma. Y entonces, se puede definir una regla específica para cuando esté incluido uno de estos verbos. Estas reglas se definieron de acuerdo a las posibles estructuras definidas por las etiquetas y cada uno de los roles de las palabras dentro de la frase de acuerdo al estándar de los GC. (ver Tabla III).

A partir de las reglas así definidas, la consecución de un grafo, corresponde a la aplicación de reglas de manera encadenada sobre el texto. Esto se puede visualizar mejor en el siguiente ejemplo:

**Frase:**

“Illustration of a neonate at autopsy whose demise was attributed to thymic death.”

**Reglas:**

- 1) Attr([Illustration], neonate)
- 2) Attr([death], thymic)
- 3) In (Attr([caption], neonate), autopsy)
- 4) Efect([demise], was attributed)
- 5) Rslt([Efect([demise], was attributed)], [Attr([death], thymic)])
- 6) Rslt(In (Attr([caption], neonate), autopsy), Rslt([Efect([demise], was attributed)], [Attr([death], thymic)]))

Por último se etiquetaron las reglas y se encontró la frecuencia de aplicación de cada regla sobre una frase. Así, en el ejemplo anterior, la regla 5 se aplicó 2 veces.

Teniendo en cuenta las reglas aplicadas en un texto, esta estará representada por un vector de frecuencias donde un valor  $f_x$  significa el número de veces que se aplicó la regla  $f$ .

TABLA III  
EJEMPLO DE REGLAS UTILIZADAS.

| Regla | Condición   | Acción o Función |
|-------|---|------------------|
| 1     | Más de un nodo del mismo concepto   | Ref(Obj,obj)     |
| 5     | (adj, obj) <b>OR</b> (adj prep obj)   | Attr(obj,adj)    |
| 5a    | (obj prep det obj) <b>OR</b> (obj prep obj) <b>OR</b> (obj prep det obj) <b>OR</b> (obj with obj) | Attr(obj,obj)    |
| 6     | (obj a_prep obj) <b>OR</b> (obj in_prep an_det obj) <b>OR</b> (obj in_prep obj)                   | In(obj,obj)      |

### B. Bigramas

Con el fin de relacionar las reglas con las etiquetas del lenguaje dentro de la oración se encontraron los bigramas de las marcas correspondientes al rol del lenguaje dentro de la oración. Se utilizaron bigramas dado que estos proveen una mayor cantidad de información al combinar las marcas consecutivas. Y la representación de cierto bigrama dentro de la frase esta dado por la presencia o ausencia del mismo.

Finalmente un texto expresado en lenguaje natural se representa por la unión de la representación de bigramas con la representación de reglas. Esto es:

*RepresentacionTexto* =  
{ *RepresentaciónBigramas*, *RepresentaciónReglas* }

### C. Función de Clasificación

Con el fin de encontrar una forma de aprendizaje sobre dichas reglas, se creó una función de clasificación basada en la sumatoria de las frecuencias de las reglas aplicadas  $\Sigma f_x$ .

## V. DESARROLLO EXPERIMENTAL

Con el objeto de obtener un mayor aprendizaje sobre las reglas, se realizaron pruebas a nivel de función o regla y a nivel de todo el conjunto de reglas.

Con base en la matriz adecuada y el software Weka [22] se utilizó un método de árbol de clasificación con el fin de aprender sobre las reglas ya previamente definidas. Los algoritmos para árboles de clasificación se basan en la cantidad de información mutua que puede darse entre una variable predictiva y su clase. Dentro de los algoritmos para árboles de clasificación más populares se encuentra el ID3 y el C4.5 [3]. Ambos métodos se basan en la cantidad de información mutua, el ID3 favorece las variables con mayor número de valores, mientras que el C4.5, corrige dicha ponderación del ID3 y realiza una poda que consiste en la aplicación de una prueba estadística para saber si se debe expandir o no una rama. El algoritmo “RandomTree” utilizado en el Software de Weka, construye un árbol partiendo de un número aleatorios de atributos para cada nodo. No realiza poda, pero utiliza un método de valoración de clase de acuerdo a las probabilidades.

En todos los casos de clasificación a la colección se le dio un tratamiento de validación por cruce con un particionamiento de 10, es decir “10 fold cross validation”

Se aplicaron los clasificadores tanto a cada regla de manera independiente como en conjunto a todas las reglas. Es decir, que se hicieron dos clases de experimentaciones, una donde la función de clase corresponde a la aplicación de una sola regla  $f$  y otra en donde la función de clase corresponde a la aplicación de todas las reglas.

### A. Resultados Experimentales por Regla

La experimentación se aplicó a cada una de las reglas deducidas. Sin embargo en el documento se incluye únicamente el análisis de dos de ellas.

1) *Regla Atributo (Attr)*: Para la regla 5 (5 y 5a) que corresponde a la aplicación de una regla *Attr* y que al aplicarla conduce según el estándar de Grafos Conceptuales al establecimiento de una relación (nodo relación de tipo Atributo) entre dos conceptos (nodos concepto). De acuerdo al árbol (ver Fig. 2) generado para esta regla se puede observar que efectivamente tiene en cuenta la aparición de un adjetivo por lo que en el nodo inicial aparece el adjetivo.

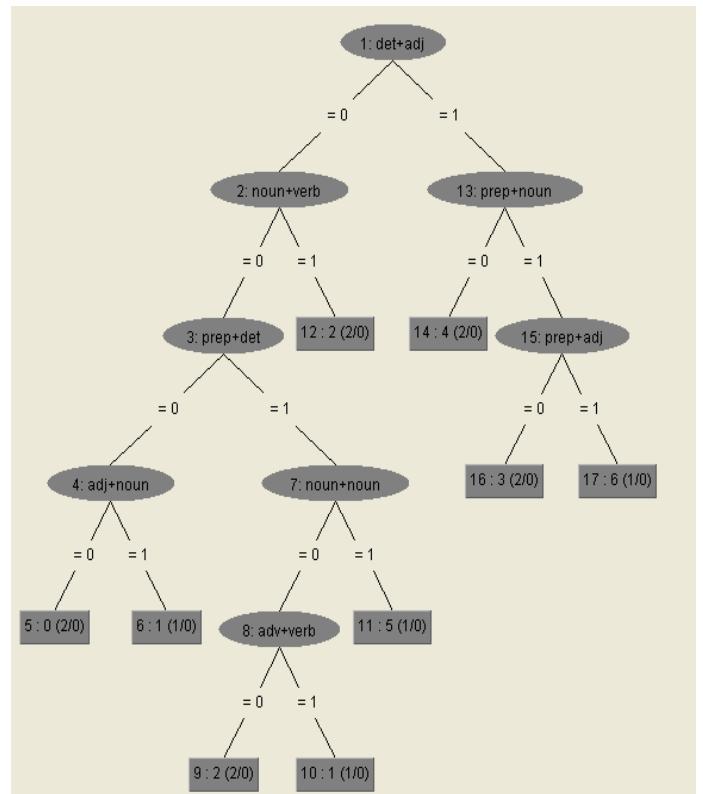


Fig. 2. Árbol generado por la regla Attr.

Con base en la (Fig 2) se pueden extractar las siguientes reglas:

a) *Regla 1*

*determinante+adjetivo+preposición+sustantivo+preposición+adjetivo.*

b) *Regla 2*

*preposición+determinante+sustantivo+sustantivo*

- c) *Regla 3*  
preposición+adjetivo
- d) *Regla 4*  
sustantivo+preposición+adjetivo
- e) *Regla 5*  
preposición+sustantivo+preposición+adjetivo
- f) *Regla 6*  
adjetivo+reposición+sustantivo+preposición+adjetivo
- g) *Regla 7*  
preposición+determinante+sustantivo

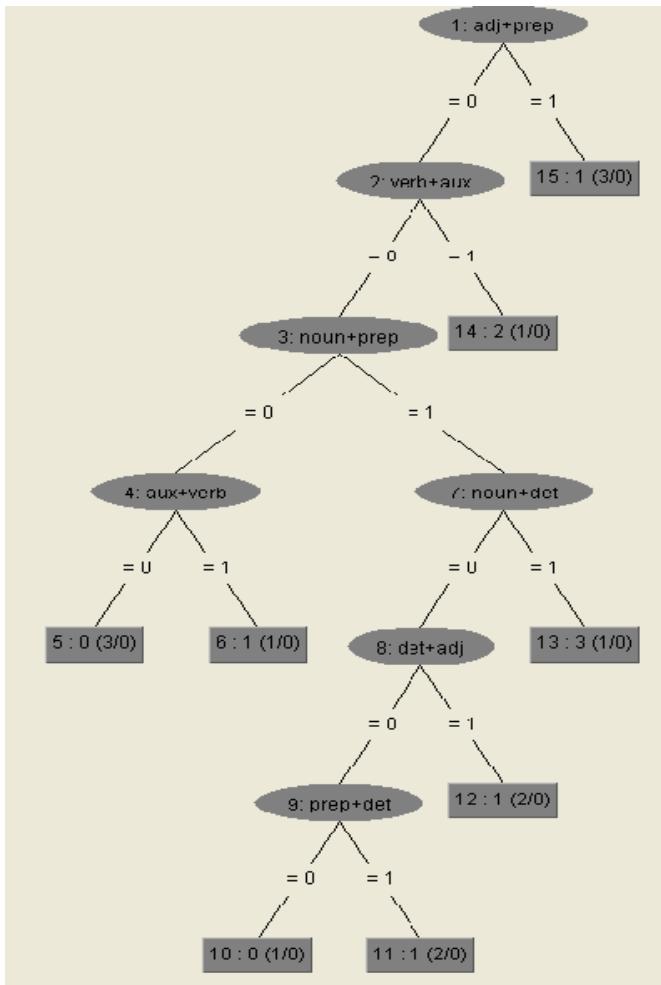


Fig. 3. Árbol Generado por la regla Rslt.

Dentro de los supuestos que se han hecho para resolver el problema está la definición del “Objeto” que se comporta como una caja negra que contiene grafos. Con el método aplicado de lo que se trata es de extractar las reglas de cada una de estas cajas. Y es así que el método ha generado nuevas reglas o ha extractado las reglas de estos objetos.

Por otro lado, al revisar las estadísticas correspondientes a la clasificación se puede observar que un 50% de los casos quedan bien clasificados.

2) *Regla Resultado (Rslt):* Esta regla define la existencia de una relación de *resultado* (*Rslt*). Un concepto es el resultado de otro, aún cuando dicho resultado no necesariamente se presenta por la aplicación de un verbo. De acuerdo al árbol

generado (ver Fig. 3) esta regla se podría aplicar en presencia de las combinaciones de:

- a) *Regla 1.*  
sustantivo+preposición+sustantivo+determinante.
- b) *Regla 2*  
sustantivo+determinante
- c) *Regla 3*  
preposición+sustantivo+determinante

De acuerdo a las estadísticas de clasificación, se obtuvo un 46% de casos bien clasificados.

De igual forma se realizó la clasificación para otras reglas obteniendo nuevas posibles combinaciones de marcaciones del idioma.

#### B. Resultados Experimentales aplicando todas las Reglas

Por último se aplicó el clasificador donde la función de clasificación corresponde a la aplicación de todas las reglas. Como resultado se obtuvieron las combinaciones que aparecen en el árbol de la Fig. 4.

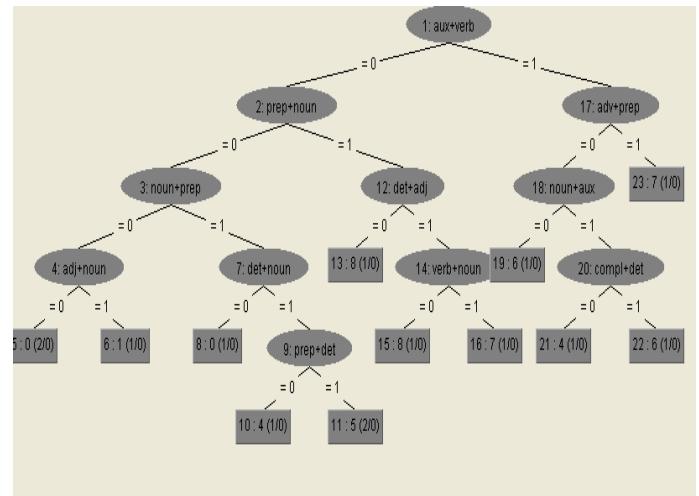


Fig. 4. Árbol Generado por la aplicación de todas las reglas.

- a) *Regla 1*  
proposición+sustantivo determinante+adjetivo+verbo+sustantivo.
- b) *Regla 2*  
sustantivo+determinante+adjetivo+verbo+sustantivo.
- c) *Regla 3*  
determinante+adjetivo+verbo+sustantivo.
- d) *Regla 4*  
adjetivo+verbo+sustantivo.
- e) *Regla 5*  
verbo+sustantivo.
- f) *Regla 6*  
sustantivo+preposición+determinante+sustantivo+preposición+determinante
- g) *Regla 8*  
preposición+determinante+sustantivo+preposición+determinante

- determinante+sustantivo+preposición+determinante*
- h) *Regla 9*  
*sustantivo+preposición+determinante*
- i) *Regla 10*  
*preposición+determinante*
- j) *Regla 11*  
*preposición+determinante+sustantivo+preposición*
- k) *Regla 12*  
*auxiliar+verbo+adverbio+preposición*
- l) *Regla 13*  
*verbo+adverbio+preposición*
- m) *Regla 14*  
*adverbio+preposición*

De acuerdo a los anteriores resultados se puede observar que efectivamente con el método propuesto si se pueden aprender nuevas reglas.

Por último se puede verificar que de acuerdo a las estadísticas del clasificador se obtuvieron resultados aceptables, con un 45% de instancias bien clasificadas.

## VI. CONCLUSIONES Y TRABAJO FUTURO

De acuerdo a los resultados obtenidos se puede concluir que se pueden deducir reglas a partir de las reglas encadenadas previamente definidas y con la definición del Objeto como artificio estructural. En un posterior trabajo se deben buscar mecanismos que permitan evaluar si la regla generada es válida para el estándar de los Grafos Conceptuales. En este sentido sería deseable hacer las pruebas necesarias para el comportamiento con trigramas e incluir técnicas que permitan tener en cuenta el orden de aplicación de las funciones.

## AGRADECIMIENTOS

Este documento se desarrolla como resultado de la estancia realizada en el Laboratorio de Lenguaje Natural del Centro Investigación en Computación del Instituto Politécnico Nacional de México. La estancia ha sido parcialmente financiada por la Universidad Nacional de Colombia y la Universidad Distrital F.J de C.

## REFERENCIAS

- [1] T. Amghar, D. Battistelli, and T. Charnois, "Reasoning on aspectual temporal information in French within conceptual graphs," in *Proc. 14th IEEE International Conf. Tools with Artificial Intelligence (ICTAI 02)*, Washington DC, 2002, pp. 315-322.
- [2] C.-Barrière and N. C. Barrière, *From a Children's First Dictionary to a Lexical Knowledge Base of Conceptual Graphs*. St. Leonard's (NSW): Macquarie Library, 1997.
- [3] Building Classification Models: ID3 and C4.5. Available: <http://www.cis.temple.edu/~ingargio/cis587/readings/id3-c45.html>.
- [4] Chih-Shyang Chang and Arbee L. P. Chen, "Supporting Conceptual and Neighborhood Queries on WWW," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: application and reviews*, Vol. 28, no. 2, pp. 300-308, 1988.
- [5] M. Hernandez Cruz, "Generador de los grafos conceptuales a partir del texto en español," Tesis de Maestría. Instituto Politécnico Nacional. Centro de Investigación en computación, 2007.
- [6] J. Farkas, "Improving the classification accuracy of automatic text processing systems using context vectors and back-propagation algorithms," in *Proc. Canadian Conference on Electrical and Computer Engineering*, University of Calgary, Alberta, 1996, pp. 696-699.
- [7] G. Allwein and J. Barwise, *Logical Reasoning with Diagrams*. New York, Oxford University Press, 1996.
- [8] A. Gelbukh and G. Sidorov, *Procesamiento automático del español con enfoque en recursos léxicos grandes*, Mexico, IPN, 2006.
- [9] A. Gelbukh and G. Sidorov, "Approach to construction of automatic morphological analysis systems for inflective languages with little effort", *Lecture Notes in Computer Science*, N 2588, Springer-Verlag, pp. 215-220, 2003.
- [10] S. Hensman, "Construction of Conceptual Graph representation of texts," in *Proc. of Student Research Workshop at HLT-NAACL*, Department of Computer Science, University College Dublin, Belfield, Dublin, 2004.
- [11] Medical Image Retrieval Challenge Evaluation. Available: <http://ir.ohsu.edu/image/2008protocol.html>.
- [12] S. Kamaruddin, A. Bakar, A. Hamdan, and F. Nor, "Conceptual graph formalism for financial text representation" in *International Symposium Information Technology*, Kuala Lumpur, Aug. 26-28, 2008, pp 1-6.
- [13] M. Last and O. Maimon, "A Compact and Accurate Model for Classification," *IEEE Trans. on Knowledge and Engineering*, Vol.16, pp. 203-215, Feb 2004.
- [14] MetaMap Transfer (MMTx). Available: <http://ii-public.nlm.nih.gov/MMTx/>.
- [15] G. W. Mineau, G. Stumme and R. Wille, "Conceptual Structures Represented by Conceptual Graphs and Formal Concept Analysis," in *Proc. of 7th Int. Conf. on Conceptual Structures (ICCS-99)*, W. Tepfenhart and W. Cyre (Eds), Springer-Verlag, 1999, pp. 423-441.
- [16] A. M. Rassinoux, R. H. Baud, C. Lovis, J. C. Wagner, and J. R. Scherrer, "Tuning Up Conceptual Graph Representation for Multilingual Natural Language Processing in Medicine Conceptual Structures: Theory, Tools, and Applications," in *Proc. 6th International Conference on Conceptual Structures, ICCS'98*, Montpellier, France, 1998, pp. 334-350.
- [17] G. Salton and M. Lesk, "The SMART automatic document retrieval systems: an illustration," *Comm. ACM*, 8 (6), pp. 391-398, Jun. 1965.
- [18] A. Schenker and H. Bunke, *Graph-Theoretic Techniques for Web Content Mining*, World Scientific Publishing, 2005.
- [19] M. Schröder, "Knowledge based analysis of radiology reports using conceptual graphs," *Lecture Notes in Computer Science*, Vol. 754 (Conceptual Structures: Theory and Implementation), Springer Berlin, pp. 293-302, 1993.
- [20] J. F. Sowa, "Semantics of conceptual graphs," in *Proceedings of the 17th Annual Meeting on Association For Computational Linguistics*, La Jolla, California, July 1979, pp. 39-44.
- [21] Unified Medical Language System (UMLS). Available: [http://www.nlm.nih.gov/research/umls/about\\_umls.html](http://www.nlm.nih.gov/research/umls/about_umls.html).
- [22] Weka software of The University of Waikato, Available: <http://www.cs.waikato.ac.nz/~ml/weka/>.
- [23] R. A. Williams, "Computational Effective Document Semantic Representation," in *Proc. of IEEE-IES Digital Eco Systems and Technologies Conf. DEST'07*, 2007.