

## **1. Revisión bibliográfica, estado del arte**

- 1.1. Lenguaje Natural Controlado: Se limita la gramática que el usuario pueda usar para hacer más sencilla a la hora de analizar. Se imponen reglas como no usar voz pasiva o evitar el uso de pronombres [1].
- 1.2. Análisis de sentimiento: Es el uso de diferentes técnicas de análisis de textos para obtener el significado subjetivo de un texto en específico.[2]
- 1.3. Búsqueda de respuestas: Consiste en la recuperación de información a partir de uno o varios texto para obtener una respuestas preguntas que se le puedan hacer.

## **2. Definición de problema, objetivos y productos.**

### **2.1. Definición del Problema**

Durante muchos años muchos científicos en el área de IA han dedicado parte de su esfuerzo en el área de reconocimiento y análisis de lenguaje natural, con el objetivo de extraer información de textos que le puedan brindar conocimiento extra a la máquina acerca de lo que está procesando. Este problema trae consigo muchas otras interrogantes, como “¿Qué se entiende por significado?” ó “¿Puede una máquina realmente entender lo que el texto quiere transmitir?”. No es conocido si se le puede dotar a una máquina de conciencia de tal modo que pueda entender y ver mundo del mismo modo que lo hace un humano, pero sí sabemos que se le puede programar para poder analizar información que pueda reinterpretar para obtener nueva información acerca de lo que el texto transmite. El presente trabajo ambiciona con dar una respuesta parcial a este problema por medio del estudio del funcionamiento de estas máquinas y aportando sugerencias de árboles de parsing para frases en español, identificando la transitividad de las oraciones e implementando reglas en el lenguaje prolog que generen redes semánticas basadas en los árboles que respondan a interrogantes de usuarios ante ciertas frases ya limitadas.

### **2.2. Objetivo General**

- Desarrollar un producto de software en Prolog que realice análisis de contenido de un texto.

### 2.3. Objetivos Específicos

- Estudiar cómo emplea Prolog el procesamiento de lenguaje natural.
- Representar árboles que analicen las categorías gramaticales aplicadas a la descomposición de sus partes.
- Implementar árboles sintácticos o árboles de parsing que representen la estructura sintagmática de oraciones de la lengua española.
- Identificar la transitividad de frases para responder a diferentes preguntas mediante la implementación de redes semánticas.

## 3. Producto

### 3.1. Definición de producto

Mediante el paradigma lógico con técnicas de producción interpretada “Prolog”, se pretende diseñar e implementar un prototipo que permita realizar procesamiento de lenguaje natural, siendo en este caso el vocabulario español. La idea es que mediante el uso de gramáticas libres de contexto y árboles de análisis sintáctico, se pueda definir la estructura correcta del lenguaje ya mencionado, caracterizando principalmente las partes que se consideren más importantes de una oración como por ejemplo: artículo, sujeto, verbo y predicado. A su vez es importante considerar el uso de gramáticas con cláusulas definidas que permitan verificar el sentido de la frase, es decir, que pueda determinar, por ejemplo, de manera correcta la oración “Los niños comen helado” y de forma incorrecta “Los comen helado”. Es importante además que las gramáticas con cláusulas definidas tomen en cuenta la concordancia de género y número, restringiendo de este modo oraciones del tipo “Los niño come helado”. Una vez teniendo definidos en el prototipo dichos términos o gramáticas es de suma importancia incorporar razonamiento de nuestro lenguaje natural, esto con el fin de lograr la meta propuesta del prototipo, que es emplear transitividad entre frases definidas, logrando que el programa sea capaz de responder o procesar interrogantes del usuario sobre una o varias oraciones ya definidas.

Primero se hará una pasada por el texto para hacer el análisis léxico de las diferentes palabras que conforman el corpus e identificar su categoría morfológica para los posteriores análisis.

Después se construirán los árboles de análisis sintáctico a partir del análisis anterior, dividiendo el texto en palabras y agrupando las palabras en los sintagmas para el posterior análisis del contenido del texto.

Por último se plantea hacer uso de redes semánticas para modelar el conocimiento que se pueda extraer del texto en específico, por medio de la red posteriormente obtener información acerca del texto que se está analizando.

### 3.2. Planificación del producto

Tabla 1: Cronograma de actividades

Fecha	Actividad	Especificación
16/04/2015	Investigación bibliográfica	Recolección de información para la definición del proyecto.
20/04/2015	Consulta con el profesor	Consulta con el profesor para delimitar el alcance y los objetivos del proyecto.
20/04/2015	Limitación del producto	Delimitación de la bibliografía y definición final del tema a desarrollar, así como la herramienta a utilizar “Prolog”.
8/06/2015	Definición del producto y conceptualización	Definición de la meta general del producto y cada una de las partes que llevan al alcance del prototipo esperado.
9/06/2015	Representación e inferencia del producto	Definir cómo se van a utilizar cada las partes definidas en la conceptualización que dará estructura a cada una de las partes del árbol de parsing hasta llegar a formar una red semántica.
9/06/2015	Casos de uso del prototipo	Dar una representación gráfica que modele la estructura lógica que deberá tener el árbol de parsing, así como una especificación de cómo será utilizado el prototipo por los usuarios finales.
14/06/2015	Creación del prototipo.	Creación del prototipo en prolog, implementando únicamente reconocimiento de oraciones que se van a utilizar.
18/06/2015	Presentación del avance	Exposición de los objetivos logrados hasta la fecha, ante el profesor y los compañeros de la clase.
22/06/2015	Corrección de los errores detectados en la presentación.	Basados en los comentarios de los compañeros y el profesor implementar mejoras en el proyecto.
02/07/2015	Implementación de redes semánticas	Como parte final del proyecto implementar redes semánticas para el análisis de la información contenida en un texto.

## 4. Conceptualización

- 4.1. Árboles de parsing: Es la representación de secuencia de derivaciones, por el cual, dada una frase, se debe descomponer en sus constituyentes sintácticos, es decir, los componentes de una frase se estructuran en diferentes categorías sintácticas. Posee nodos internos que representan los símbolos no terminales y hojas que constituyen los símbolos terminales. En base al lenguaje natural estos símbolos terminales y no terminales, se encontrarán representados de acuerdo a la sintaxis composicional: sintagma nominal y sintagma verbal.
- 4.2. Oración: Regla de inicialización, conformada por una frase que puede ser descompuesta o concatenada por dos cláusulas distintas del tipo sintagma nominal y sintagma verbal.
- 4.3. Sintagma nominal: Constituyente sintáctico dada por la concatenación de cláusulas de tipo artículo y sujeto, donde el sujeto puede ser un sustantivo o pronombre.
- 4.4. Sintagma verbal: Regla que procesa cláusulas de tipo verbo y predicado donde el predicado puede corresponder también un sujeto o pronombre.
- 4.5. Artículo: Categoría morfológica que da conocimiento y concretividad al sustantivo o pronombre.
- 4.6. Sujeto: Sintagma que será participante de la predicación verbal, puede designar un objeto, persona, animal o cosa abstracta
- 4.7. Verbo: Sintagma léxico que representa la acción realizada por los constituyentes del sintagma nominal.
- 4.8. Complemento: Constituyente del sintagma verbal que da descripción de la acción realizada por el sujeto, puede describir el estado en el que se encuentran las cosas en de los constituyentes de la oración.
- 4.9. Red semantica:
- 4.10. Token:
- 4.11. Tipos:
- 4.12. Nodos:
- 4.13. Arcos:

Tabla 2: Relaciones conceptuales

	Artículo	Sujeto	Verbo	Complemento
Oración	✓	✓	✓	✓
Sintagma Nominal	✓	✓		
Sintagma Verbal			✓	✓

## 5. Representación e inferencia

Definición de gramáticas:

La gramática del lenguaje natural será definida inicialmente por una notación de gramáticas que serán establecidas mediante un árbol de parsing con el paradigma Prolog, el objetivo primordial es establecer un conjunto de reglas de inferencia que ayuden a generar cláusulas o listas debidamente definidas por nombres y especificaciones. En base a la conceptualización de la sección anterior se define la estructura preliminar que tendrá el árbol de análisis sintáctico (ver figura 1).

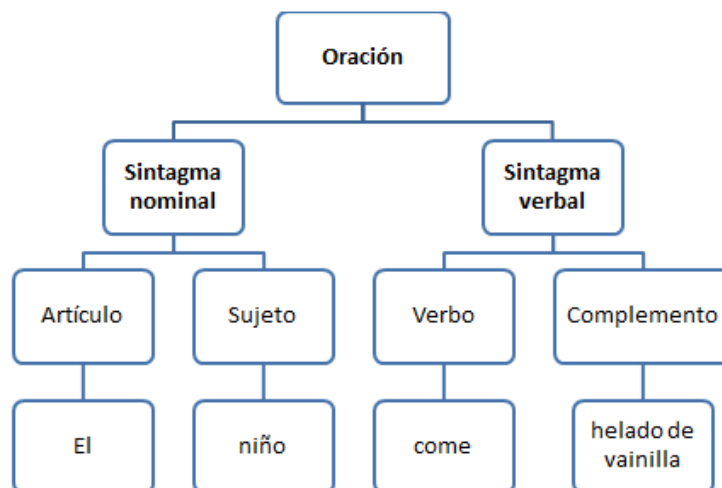


Figura 1: Árbol de parsing

//operadores and entre cada subárbol, operador or para determinar si es un artículo and nombre o predicado

#### Análisis de Gramáticas:

Mediante una red semántica se dará una muestra del razonamiento del lenguaje en la que los conceptos y sus interrelaciones tendrán como objetivo dar respuestas a interrogantes en base a una frase previamente definida con los árboles de parsing (ver figura 2).

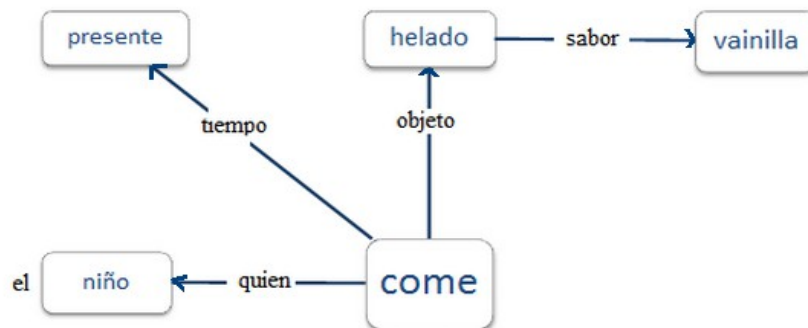


Figura 2: Redes Semánticas

#### 6. Casos de uso

#### 7. Marco teórico //esto no es parte de este documento pero hay que considerar

diferencia de listas

#### 8. Desarrollo

#### 9. Validación y Plan futuro

## 10. Referencias Bibliográficas

[1] S. (2003). *Controlling Controlled English An Analysis of Several Controlled Language Rule Sets* Consultado el 30 de Abril del 2015, de <http://www.mt-archive.info/CLT-2003-Obrien.pdf>

[2] Pang, B. & Lee, L. (2002). *Thumbs up? Sentiment Classification using Machine Learning Techniques* Recuperado el 15 de Junio del 2015, de <http://www.cs.cornell.edu/home/llee/papers/sentiment.pdf>

[3] L. HIRSCHMAN and R. GAIZAUSKAS (2001). Natural language question answering: the view from here. *Natural Language Engineering*, 7, pp 275-300. doi:10.1017/S1351324901002807.

[4] Aho, A. (2008). *Compiladores: Principios, técnicas y herramientas*. México : Pearson.

[5] Bratko, I. *Prolog Programming for Artificial Intelligence: Cap 21: "Language Processing with Grammar Rules"* Pearson.

[6] L. HIRSCHMAN and R. GAIZAUSKAS (2001). Natural language question answering: the view from here. *Natural Language Engineering*, 7, pp 275-300. doi:10.1017/S1351324901002807.

[https://www.cs.unm.edu/~luger/ai-final2/CH8\\_Natural%20Language%20Processing%20in%20Prolog.pdf](https://www.cs.unm.edu/~luger/ai-final2/CH8_Natural%20Language%20Processing%20in%20Prolog.pdf)

<http://www.mtome.com/Publications/PNLA/prolog-digital.pdf>