

Acceso a un Sistema de Inventarios usando Procesamiento de Lenguaje Natural y RIAs

J. Ferrer D., C. C. Ríos, M. G. Sandoval, R. Baltazar, J. M. Carpio, M. Ornelas

Instituto Tecnológico de León (México)

Departamento de Sistemas y Computación. Av. Tecnológico s/n Fracc. Julián de León, Gto., C.P. 37290

Tel.: +(477)7105200, ext. 5051, E-mail: jazfer_1688@hotmail.com, karo_3108@hotmail.com,

dear_lupita@hotmail.com, charobalmx@yahoo.com.mx, jmcarpio61@hotmail.com,

mornelas67@yahoo.com.mx,

Resumen — En este documento se detalla un sistema desarrollado para interpretar lenguaje natural. Dicho sistema, tiene como objetivo principal, la obtención de datos a través de tratamiento del texto, destacando las siguientes tareas de clasificación: la *resolución de la ambigüedad léxica* y el *etiquetado sintáctico*. Además se reduce la brecha existente entre tareas técnicas computacionales y ordenes simples en el lenguaje cotidiano. El Sistema de Procesamiento de Lenguaje Natural fue desarrollado en ZK con plataforma JAVA, en combinación con otra herramienta que lo complementara es el Navicat for MySQL para la realización de la base de datos y el diccionario de palabras. La necesidad de un sistema de este tipo, es grande, pues cuando se accesa a un sistema con implementación de base de datos la gran cantidad de usuarios desconoce la estructura léxica y sintáctica de los sistemas informáticos.

Abstract – In this document is presented detailed information about software developed to translate or interpret natural language. The main purpose of this system is to get data through text treatment, highlighting the following classify task: the resolution of lexica ambiguity and the syntactic labeling. In addition this software bridges the computational techniques task and simple orders on daily language. The system for natural language processing was developed on ZK with JAVA platform, combined with other tools that will complete the development such as Navicat for

MySQL that helps with the data base and the word dictionary. The necessity of this kind of system is huge, because the majority of the users of computational services are not aware about the lexica and syntactic structure of the information systems.

Descriptorios – Inteligencia Artificial, Procesamiento de Lenguaje Natural, Sistema de inventarios.

I. INTRODUCCIÓN

Una de las ramas más importantes de la Inteligencia Artificial es aquella orientada a facilitar la comunicación hombre-computadora por medio del lenguaje humano, o lenguaje natural. El Procesamiento del Lenguaje Natural (PLN) es la disciplina encargada de producir sistemas informáticos que posibiliten dicha comunicación, por medio de la voz o del texto. Existen varios trabajos que tratan el problema del Procesamiento del Lenguaje Natural [1-2], y en el caso específico de las bases de datos y su acceso mediante el PLN, lo trata Rocha et al [3] usando redes neuronales. En este trabajo se realiza el procesamiento para el Idioma Español usando Análisis léxico, sintáctico y semántico. Se construyó un lexicón apropiado para una base de datos de inventarios de equipos.

El *lenguaje natural* es el lenguaje hablado o escrito por humanos para propósitos generales de comunicación, es esencial para la comunicación entre

humanos y máquinas, el desarrollo de esta área permite la posibilidad de construir traductores, interfaces a sistemas de información, etc. Desde la perspectiva de la inteligencia artificial (IA), el estudio del lenguaje natural tiene dos objetivos:

1. Facilitar la comunicación con la computadora para que accedan a ella usuarios no especializados.
2. Modelar los procesos cognoscitivos que entran en juego en la comprensión del lenguaje para diseñar sistemas que realicen tareas lingüísticas complejas (traducción, resúmenes de textos, recuperación de información, etc.)

El PLN se concibe como el reconocimiento y utilización de la información expresada en lenguaje humano a través del uso de sistemas informáticos. En su estudio intervienen diferentes disciplinas tales como lingüística, ingeniería informática, filosofía, matemáticas y psicología. Debido a las diferentes áreas del conocimiento que participan, la aproximación al lenguaje en esta perspectiva es también estudiada desde la llamada ciencia cognitiva [4].

A. Antecedentes

En 1970 Halstead analizó prosa en inglés utilizando smalltalk. Con este trabajo, se definió una correspondencia entre el lenguaje natural y las primitivas computacionales.

En 1980, Abbott propuso un enfoque para el diseño del programa Ada basado en el análisis lingüístico de las estrategias informales escrito en inglés. Booch amplió este enfoque en un objeto de forma explícita el proceso de diseño orientado a objetos.

En 1989, Saeki, Horai, y Enomoto propusieron un proceso de diseño de software basado en lenguaje natural. Su trabajo explica en detalle y complementa la heurística ofrecido por Abbott. Se centraron sobre todo en la identificación del comportamiento del sistema dinámico como el expresado por los verbos en una descripción en lenguaje natural.

En 1997, Adam and Gangopadhyay propusieron una metodología para desarrollar una interfaz que combina interfaces de SQL con un sistema de manejo de bases de datos usando lenguaje natural.

B. Análisis del Procesamiento del Lenguaje Natural

Las principales áreas de trabajo en el PLN son: Síntesis del discurso, Análisis del lenguaje, Comprensión del lenguaje, Generación de lenguajes naturales, Traducción automática, Respuesta a preguntas, Recuperación de la información, Extracción de la información.

En PLN se tienen varios niveles para realizar el análisis de la información expresada en lenguaje humano y son los siguientes:

- *Fonética*. Formación de palabras a partir de sonidos.
- *Análisis Morfológico*. Análisis de las palabras para extraer raíces, rasgos flexivos, unidades léxicas compuestas y otros fenómenos.
- *Análisis Sintáctico*. El análisis de la estructura sintáctica de la frase mediante una gramática de la lengua en cuestión.
- *Análisis Semántico*. La extracción del significado de la frase, y la resolución de ambigüedades léxicas y estructurales.
- *Análisis Pragmático*. El análisis del texto más allá de los límites de la frase, por ejemplo, para determinar los antecedentes referenciales de los pronombres.

El desarrollo de software en general siempre comienza con una necesidad como la de explorar o resolver un problema, entretener, educar o compartir la información y el conocimiento. Al utilizar el lenguaje natural en software los problemas a menudo son complejos, vagos y ambiguos. Este tipo de problemas pueden ser complejos cuando contienen cláusulas y frases que describen y relacionan varios objetos, condiciones, eventos y/o acciones. Vagos, es decir cuando contienen generalizaciones, o les falta información importante, también pueden ser ambiguos cuando son abiertos a múltiples interpretaciones. Esto nos lleva a la necesidad de realizar un sistema que evite estos problemas, ya que, es cada vez más grande el número de usuarios de computadoras que no tienen un conocimiento profundo de los comandos u órdenes internas requeridas para un mejor aprovechamiento de los recursos de información que provee un sistema computacional, por este motivo se inició la investigación sobre el procesamiento de lenguaje natural. Se consideró el desarrollo de un sistema capaz de interpretar consultas en lenguaje natural sobre un sistema de inventario, que comprenda instrucciones

en lenguajes de más alto nivel, como el lenguaje natural o de uso cotidiano, en un dominio restringido.

El objetivo del proyecto presentado en este artículo es obtener información de manera fácil y eficiente, pues teniendo una interfaz en lenguaje natural, los usuarios no necesitan conocer secuencias de comandos o instrucciones específicas de un lenguaje. Con este sistema cualquier usuario que tenga un conocimiento básico sobre las características del sistema de información es capaz de interactuar con el sistema de inventarios. Nótese que la etapa actual del sistema no permite instrucciones en lenguaje natural muy complejas.

II. DESCRIPCIÓN DEL SISTEMA

El proyecto consta de un sistema de inventarios que permitirá llevar el control del equipo de cómputo, impresoras, mobiliario y demás. El procesamiento del lenguaje natural será aplicado en las consultas que se harán al sistema. De las consultas se desglosarán informes y gráficas. El proyecto será desarrollado en ZK y MySQL[5] para la base de datos del sistema y del diccionario de palabras. Para resolver los requerimientos y además ofrecer un sistema independiente de la plataforma, el sistema será desarrollado utilizando una interfaz vía Web.

Por lo que se aprecian dos actividades principales:

1. *Creación del sub-módulo para procesar el Lenguaje Natural:* Donde se encarga del procesamiento de una oración de lenguaje natural, efectuando el análisis léxico y sintáctico para formar las acciones correspondientes que se requieran para la realización de la consulta. Fue implementado el lenguaje JAVA y como manejador de la base de datos se utilizó Navicat for MySQL.
2. *Desarrollo del sub-módulo de interfaz web RIA (Aplicación Enriquecida de Internet):* Su tarea principal es brindar una interfaz amigable entre el módulo de procesamiento de lenguaje natural y el usuario. Además, proporcionar un medio eficaz para que el usuario incremente las posibilidades del sistema al permitirle introducir más platillas de oraciones, palabras y acciones.

C. Fases del Proyecto

Para el correcto desarrollo del presente proyecto se implementaron 3 tipos de analizadores: el analizador léxico, en el cual se recibe como entrada una secuencia de caracteres y se produce una salida de Tokens (componentes léxicos o símbolos), el analizador sintáctico, con este analizador se reconoce si una o varias cadenas de caracteres forman parte de un determinado lenguaje y posteriormente se convierte el texto de entrada en otras estructuras (comúnmente árboles), que son más útiles para el posterior análisis y capturan la jerarquía implícita de la entrada. Finalmente el analizador semántico se encarga de recibir la frase validada por el analizador sintáctico para interpretarla obteniendo la información requerida. A continuación se muestra cada uno de los analizadores.

Analizador Léxico

Para desarrollar la fase del programa que reconozca las ocurrencias de palabras, nuestro primer paso sería representar de una manera concisa, no ambigua la estructura de una palabra aceptable. Para este fin utilizamos la forma gráfica de un *diagrama de transiciones* (también llamado *diagrama de estado* o, en el campo del procesamiento de lenguajes naturales, *red de transiciones*). (Figura 1).

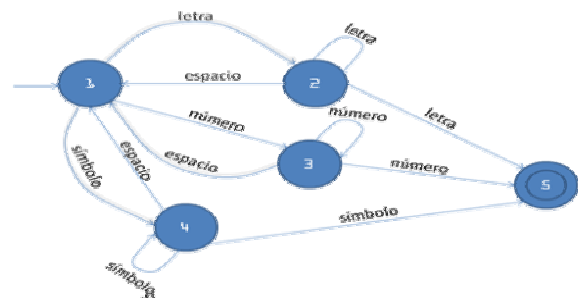


Figura 1. Autómata del Analizador Léxico. Para reconocer y separar las palabras, inicia en estado 1 y continua al estado 2 si el siguiente carácter es una letra, siguiendo así hasta terminar con las letras de la palabra, igual para los números y serie de símbolos.

Analizador Sintáctico

Para desarrollar el analizador sintáctico lo primero es establecer una gramática independiente del contexto para las frases que serán aceptadas por el sistema.

Después se verificará que las palabras o tokens obtenidos del analizador léxico, están dentro del diccionario para el lenguaje. Posteriormente se verificara con árboles de decisión que se cumple con la gramática propuesta [6-9]. Figuras 2-5.

Gramática Principal

- Frase → Verbo Cadena
- Verbo → mostrar | ver | generar | ...
- Cadena → Artículo Sustantivo | Complemento |
- Artículo Complemento | Artículo Adjetivo | Acciones
- Adjetivo | Sustantivo
- Artículo → el | las | los | la | ...
- Sustantivo → monitor | importe | equipo | ...
- Adjetivo → Preposicion Termino
- Preposicion → a | ante | bajo | ...
- Termino → nombreProyecto | clave | nombreResponsable | ...
- Acciones → ingresado | capturado | ...
- Complemento → Sustantivo Del

Gramática de Fecha

- Termino → fechaPeriodo | Fecha1
- fechaPeriodo → Fecha1 Separador Fecha1
- Separador → al | - | a
- Fecha1 → Num De Mes Del Año | Mes Num Del Año | Mes Num Separador1 Año | Mes Num Separador2 Año | Mes Del Año | Año Separador1 Mes Separador1 Año | Num |Año | Num Separador1 Mes Separador1 Año | Mes |Num Separador2 Mes Separador2 Año | Num De Mes | Mes Num
- Num → 01 | 02 | 03 | ...
- De → de
- Del → del
- Separador1 → /
- Mes → 01 | 02 | ... | 12 | Enero | Febrero | ... | Año → 2000 | 2001 | 2002 | ... | 2009

Árboles del Analizador Sintáctico

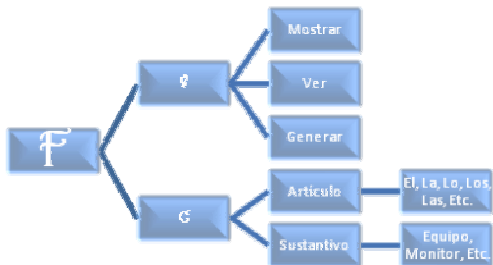


Figura 2. Árbol del analizador sintáctico que demuestra la transición que sigue la gramática para una frase de la forma “mostrar los equipos”.

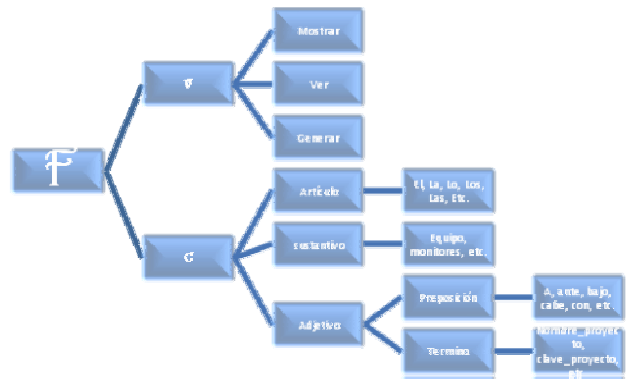


Figura 3. Árbol del analizador sintáctico que demuestra la transición que sigue la gramática para una frase de la forma “mostrar los equipos del proyecto x”.

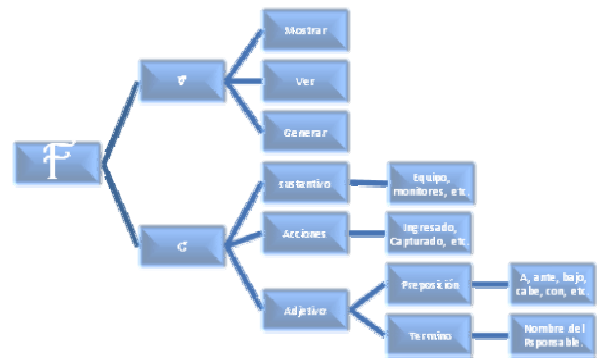


Figura 4. Árbol del analizador sintáctico que demuestra la transición que sigue la gramática para una frase de la forma “mostrar impresoras de Administrador”.

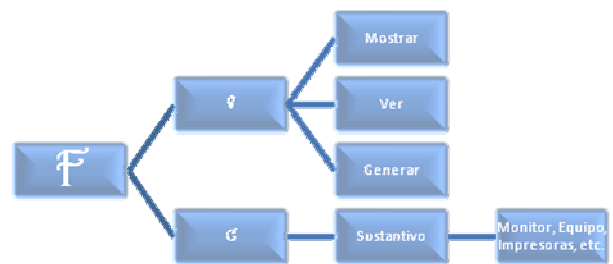


Figura 5. Árbol del analizador sintáctico que demuestra la transición que sigue la gramática para una frase de la forma “ver equipos”.

Frases aceptadas por la Gramática:

- Mostrar monitores.
- Ver equipos de fecha1 a fecha2.
- Generar importe de fecha1 a fecha2.
- Mostrar equipos de Responsable (nombre).
- Mostrar equipos de Nombre de proyecto.
- Mostrar equipos de Clave de proyecto.
- Mostrar equipos Ingresados en fecha.
- Mostrar equipos con números de inventario.

El sistema es capaz de identificar una frase escrita como naturalmente se preguntaría a una persona encargada de un almacén o del inventario de una institución como se muestra en la siguiente frase: “Mostrar los equipos del Proyecto de Fomix” y el sistema mostrará los equipos clasificados en el proyecto Fomix, es decir no necesitará de elementos como botones o menús para realizar la consulta.

III. CONCLUSIÓN

En este proyecto se ha presentado una técnica para el procesamiento del lenguaje natural empleado en una aplicación enriquecida de internet para un sistema de inventarios. El empleo del lenguaje natural en este proyecto permite al usuario obtener información específica, más allá de lo que esté predefinido en el sistema, esto ha mostrado gran flexibilidad, ya que las frases aplicadas han sido reconocidas con éxito y, por consiguiente, los informes requeridos correspondientes a las frases capturadas en dicho sistema de inventarios se han efectuado y los resultados de esto son mostrados al usuario; teniendo como producto un sistema que puede crecer de acuerdo a diferentes necesidades.

REFERENCIAS

- [1] Minhwa, C., Moldovan, D., “Parallel Natural Language Processing on a Semantic Network Array Processor”, *IEEE Transactions on Neural Networks*, Vol. 3 No. 5 September 1992, pp. 819-828.
- [2] Peter, C., Henderson, J., “Incremental Syntactic Parsing of Natural Language Corpora with Simple Synchrony Networks”, *Transactions on Knowledge and Data Engineering*, Vol. 13 No.2, March-April 2001, pp. 219-231.
- [3] Adam, N., “A Form-Based Natural Language Front-End to a CIM Database,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 9 No.2, March-April 2007, pp. 238-250.

- [4] Allen, J., *Natural Language Understanding*, Addison Wesley, 1994.
- [5] WWW.POSTGRES.ORG
- [6] Gelbukh, A., *Using a semantic network for lexical and syntactic disambiguation*, Noviembre 1997.
- [7] Gelbukh, A., Bolshakov, I., Galicia-Haro, S., “Statistics of parsing errors can help syntactic disambiguation”, *CIC-98 - Simposium Internacional de Computación*, Noviembre 11 - 13, 1998.
- [8] Bolshakov, I., Gelbukh, A., Galicia-Haro, S., “Syntactical managing patterns for the most common Spanish verbs,” *Simposium Internacional de Computación* 1997.
- [9] Bolshakov, I., Gelbukh, A., Galicia Haro, S., Orozco Guzmán, M., “Government patterns of 670 Spanish verbs”, *CIC-98 - Simposium Internacional de Computación*, Noviembre 11 - 13, 1998.