

Procesamiento del lenguaje natural basado en una “gramática de estilos” para el idioma español.

Hilda Y. Contreras Z.
Postgrado en Computación.
Universidad de Los Andes.
Mérida, Venezuela.
hylitza@yahoo.com

Jacinto A. Dávila Q.
Centro de Investigación y Proyectos en Simulación y Modelos, Postgrado en Computación.
Universidad de Los Andes.
Mérida, Venezuela.
jacinto@ula.ve

RESUMEN

Este artículo presenta un proyecto de investigación que pretende desarrollar una herramienta para interpretar documentos en español y extraer de ellos descriptores relevantes. Los problemas de procesar el lenguaje natural y de extraer información, han sido atacados desde hace varias décadas [13], [2], [18]. Sin embargo, las investigaciones no han sido suficientes para diseñar un sistema que interprete el lenguaje natural con un rendimiento cercano al de un humano. El lenguaje natural escapa a todos los esfuerzos de tratamiento computacional, al parecer, debido a que el conocimiento lingüístico está asociado de formas sutiles y desconocidas con el conocimiento contextual que tiene el hablante [15]. En este trabajo abordaremos el problema de la interpretación del lenguaje escrito usando gramáticas de estilos y formas lógicas. La gramática de estilo se inspira en las reglas de estilo que propone *J. Williams* [16] para escribir prosa en inglés. Esta estrategia adaptada al español y la definición de un buen descriptor, tienen la finalidad de reducir la complejidad del procesamiento sintáctico/semántico tradicional; Además de incorporar el conocimiento contextual en el proceso. Validaremos la estrategia con un prototipo de un módulo de asignación de descriptores para un sistema bibliográfico virtual.

Palabras claves: Procesamiento del lenguaje natural, Lingüística computacional, Minería de datos, Recuperación de información, Inteligencia artificial.

Natural language processing based on “grammars of style” for Spanish language.

ABSTRACT

This paper presents a research project which tries to develop a system to process documents in Spanish and extract good descriptors from them. Natural languages have been the subject of numerous attempts of systematic analysis and the development of automatic, sense-extraction mechanisms [13], [2], [18]. However, all this research effort has been unable, as yet, to develop a system with human-like performance in natural language processing. It seems that natural languages escape from computational treatment due to the fact that linguistic knowledge is intertwined with some very subtle and some unknown forms of contextual knowledge easily reachable by human beings [15]. This work will concentrate efforts on written natural language processing by means of grammars of style and logic-based knowledge representation. The grammars of style that are being used are based on those proposed by Williams [16] for the English language. The style-based strategy, its adaptation to Spanish and logical definitions for good descriptors should reduce the complexity of the syntactical/semantical analysis, making the project feasible.

Keywords: Natural language processing, Computational linguistics, Data mining, Information retrieval, Artificial intelligence.