

Informe de Oponencia

Tesis: Generación Automática de Ontologías

Autor: José Ariel Romero Costa

Oponente: Suilan Estévez Velarde

La representación computacional del conocimiento es una de las tareas más importantes en la construcción de sistemas inteligentes capaces de manejar automáticamente grandes volúmenes de datos y tomar decisiones basadas en estos. Dentro de esta problemática, las ontologías constituyen uno de los mecanismos más usados para almacenar de forma computacionalmente conveniente la formalización del conocimiento acerca de un dominio concreto. Debido a su propia naturaleza, la construcción manual de ontologías es una tarea costosa, por lo que la generación automática o semi-automática de estas a partir de contenido no estructurado, por ejemplo, lenguaje natural, es un área de investigación muy activa.

En esta problemática se desarrolla la tesis de José Ariel, con una propuesta para la construcción automática de una base de conocimientos a partir de la anotación semántica de oraciones en lenguaje natural. Partiendo de un esquema de anotación genérico capaz de capturar la semántica más relevante de una oración factual, su propuesta consiste en un proceso de normalización y agregación de las entidades y relaciones anotadas a nivel textual para construir un grafo que contenga el conocimiento del corpus analizado.

Todo el esfuerzo realizado y los resultados obtenidos son recogidos en un documento que resume, en Introducción y 4 capítulos centrales, los principales retos e ideas de la investigación. En el estado del arte se abordan los conceptos fundamentales relacionados con la representación de conocimiento, las principales estrategias para la generación automática de ontologías y su evaluación. En el capítulo 2 se describe el modelo de anotación utilizado. El capítulo 3 describe el proceso utilizado para construir un grafo de conocimiento a partir de las anotaciones semánticas. Se propone una estrategia inicial adecuada, pero sería interesante explorar técnicas más avanzadas para mejorar la calidad del conocimiento obtenido. El capítulo 4 presenta un marco experimental donde se brindan estadísticas sobre los resultados obtenidos, aunque sería deseable una evaluación cualitativa que permita comprobar la calidad del conocimiento extraído.

Esta investigación fue desarrollada en condiciones no ideales debido a la situación epidemiológica, lo cual ha impactado negativamente en los resultados experimentales. Aún así, la tesis demuestra una adecuada aplicación del método científico dentro de las limitaciones existentes. En función de estas consideraciones, se plantean las siguientes preguntas para profundizar en las cuestiones menos discutidas en la tesis:

1. Realice una comparación entre su metodología para construir el grafo de conocimiento y los principales sistemas de *ontology learning* existentes.
2. Realice una evaluación cualitativa, por expertos, de la ontología obtenida a partir de todo el corpus, y evalúe la calidad del conocimiento generado.
3. Proponga una estrategia para resolver el problema de correferencias.
4. Dado que en una ontología el contenido aparece usualmente sintetizado o resumido, ¿puede proponer una estrategia, aplicable a su metodología, para resumir contenido?



M.Sc. Suilan Estévez Velarde