

ESTUDIO, DEFINICIÓN E IMPLEMENTACIÓN DE MÉTRICAS DE SIMILITUD ENTRE FRASES

**MARCOS ANTONIO GUTIÉRREZ SOTOMAYOR
INGENIERO CIVIL EN COMPUTACIÓN**

RESUMEN

La refinación de las traducciones automáticas pretende corregir ciertas ambigüedades que ocurren con los algoritmos que actualmente se usan en los motores de traducción, de modo de obtener textos cuya comprensión por parte del usuario final sea mejor. Nuestra propuesta se basa en tomar como fuente de conocimiento la Web, entorno donde en cada momento se esta publicando nueva información (revistas, libros, tesis, papers, investigaciones, noticias etc.) en diferentes lugares del mundo. En este estudio se considera que la unidad minima de traducción es una frase. La máquina computacional considerada para llevar acabo el proceso de refinamiento considera tres etapas a saber: (1) la captura de información textual de la Web (que deja como resultado una base de datos que crece constantemente); (2) El indexamiento de la información recopilada; y posteriormente, (3) algoritmos de búsqueda y métricas por similitud que permiten encontrar la frase mas adecuada al contenido (contexto) del texto traducido. Esta memoria esta enfocada al estudio, definición e implementación de métricas de similitud entre frases, con las cuales se pueda medir que tan semejante es una frase con respecto a otra. Para esto se estudiarían diferentes criterios de similitud entre frases, se desarrollaran técnicas para recuperar información relevante que logre ser filtrada por los criterios de similitud ya estudiados, y por último de implementaran métricas que logran definir que tan semejantes son una frase de otras. Palabras Clave: Criterios de Similitud, Recuperación de la Información, Métricas de Similitud entre Frases.