

RECUPERACIÓN EFICIENTE DE DATOS PARA EL REFINAMIENTO DE TRADUCCIONES AUTOMÁTICAS, UTILIZANDO LA WEB COMO BASE DE CONOCIMIENTO

**DANILO ALFONSO YÁÑEZ ALARCÓN
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 mínima de traducción es una frase. La maquina 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 por símil aridad que permiten encontrar la frase más adecuada al contenido (contexto) del texto traducido. Esta memoria se enfoca principalmente en el indexamiento de frases y la resolución eficiente de diversos tipos de consultas para recuperación de frases utilizando diferentes estructuras de datos y algoritmos que servirían de apoyo al momento de tomar la decisión de las frases mas apropiadas a fin de mejorar la calidad de la traducción. Esta memoria esta inmersa en un proyecto mas complejo en que se estudian y proponen técnicas que permite refinar las traducciones automáticas, previamente realizadas por motores de traducción al idioma español, por ejemplo: *Google Translate, Yahoo! Babel Fish, Systran* entre otros. Dicha memoria se enfoca principalmente en la recuperación eficiente de información textual como medio de apoyo (reflejada en un modulo de la aplicación) al proceso de refinamiento de traducciones automáticas, indexando y estructurando la información que ha sido obtenida de la Web. Palabras Clave: Recuperación de Información, Indexación, Índices invertidos, Lenguajes de consulta.

ABSTRACT

Automatic translation refining aims to correct some ambiguities that occur when using algorithms that are currently operating on translation engines, so as to obtain texts that are more understandable by the end user. Our proposal is based on taking the Web as a knowledge source where every time new information (magazines, books, theses, papers, research, news, and so on) is being published in different parts of the world. In this study we consider that the minimum unit of translation is a phrase. The whole machinery required to carry out the refining process considers three stages, namely, (1) to capture textual information from the Web (this stage produces a phrase database that is constantly growing); (2) to index the information collected; and (3) similarity search algorithms that find the most appropriate phrase according to the content (context) of the translated text. In this study we consider that the minimum unit of translation is a phrase. While the whole project studies and proposes techniques that allow to refine automatic translations previously translated by a engine translations (for instance, *Google Translate*, *Yahoo! Babel Fish*, *Systran*) into Spanish; this work focuses primarily on the efficient retrieval of textual information as a mean of supporting the process of refining machine translations, indexing and structuring the information that has been obtained from the Web.