



**Universidad de Concepción - Chile**  
**Facultad de Ingeniería**  
**Departamento de Informática y Ciencias de la Computación**

# **Geo-Referenciación con Adquisición Semi-Automática de Conocimiento**

**Julio Erasmo Godoy del Campo**

**Profesores Guía**

**María Andrea Rodríguez Tastets**

**John Atkinson Abutridy**

**Tesis presentada a la**  
**ESCUELA DE GRADUADOS**  
**DE LA UNIVERSIDAD DE CONCEPCIÓN**

**Para optar al Grado de**  
**MAGÍSTER EN CIENCIAS DE LA COMPUTACIÓN.**

**Abril de 2008**

# Resumen

La World Wide Web se compone de distintos tipos de información, y su cantidad crece de manera vertiginosa. Inicialmente los buscadores operaban de una misma forma para todos los tipos de consultas realizadas. Sin embargo, en el contexto de la información geográfica, los métodos tradicionales de búsqueda resultan ineficientes, y se requiere el desarrollo de metodologías específicas para este dominio. Los avances en este tema se basan principalmente en la utilización de una Gazetteer y estrategias de geo-referenciación de documentos Web, lo que permite mejorar el proceso de búsqueda usando criterios geográficos. Estos métodos, sin embargo, se caracterizan por estar limitados a usar un conjunto de lugares geográficos predefinido y no consideran otros aspectos propios del dominio, en particular, la relación espacial entre localidades geográficas que pudieran asociar una geo-referencia a un documento Web, la cual no haya sido explícitamente citada en el texto del documento.

En esta tesis se propone un modelo descriptivo de un sistema que permita obtener el conjunto de lugares geográficos a partir de documentos Web, considerando las relaciones espaciales entre localidades geográficas para una mejor precisión en la geo-referenciación de documentos y la incorporación dinámica de nuevos elementos a una Gazetteer. A diferencia de trabajos anteriores, esta tesis combina técnicas de recuperación de información tradicional con técnicas de procesamiento de lenguaje natural. El resultado de esta tesis se implementa en un prototipo del modelo propuesto, cuyos resultados demuestran las ventajas de incorporar las metodologías propuestas en la detección y asociación de lugares geográficos a documentos de texto.