



**Universidad de Concepción
Facultad de Ingeniería
Departamento de Ingeniería Civil Informática
Programa Magister en Ciencias de la Computación**

Clasificación Adaptiva para la Identificación de Información Retórica de Textos en Lenguaje Natural

Elvis Cristián Aravena Beltrán



Profesor Guía

John Atkinson Abutridy

**Tesis de Grado Presentada a la
ESCUELA DE GRADUADOS
DE LA UNIVERSIDAD DE CONCEPCION**

**Para Optar al Grado de
MAGÍSTER EN CIENCIAS DE LA COMPUTACIÓN**

**Concepción – Chile
Septiembre, 2007**

Resumen

Al procesar grandes cantidades de textos electrónicos para el propósito de análisis y minería de datos, existen diferentes niveles de conocimiento. Asumiendo que la tarea de análisis sintáctico ya puede ser resuelta utilizando técnicas de análisis sintáctico parcial, el siguiente paso es intentar capturar conocimiento a nivel de las relaciones retóricas (conocimiento a nivel de discurso) de dichos documentos, lo que es de vital importancia para aplicaciones como generación automática de resúmenes, minería de textos y extracción de información.

En este contexto, el proceso de identificación y extracción de información a nivel de discurso puede pensarse como un proceso de aprendizaje supervisado a partir del cual un modelo aprende a identificar determinadas relaciones retóricas o de discurso (ej. determinar dónde, en un texto, un autor hace referencia a los resultados del trabajo, su background, conclusiones, etc., y cómo éstos están conectados). El objetivo del trabajo es, entonces, desarrollar un modelo de clasificación en base a técnicas de aprendizaje computacional (ej. Algoritmos Genéticos) y de técnicas de extracción de información, que permitan la captura de conocimiento retórico de textos en lenguaje natural en Español.

