



Tesis: “Análisis del algoritmo ActiveRank como método de detección automática de contenido dentro de redes de información”

Fernando Luege Mateos

México D.F., Febrero 2010



## **Prefacio**

La capacidad con la que contamos hoy en día para generar y compartir información, a través de medios digitales y redes globales, es simplemente ilimitada; de igual manera, los retos de ingeniería que presentan la recopilación, el análisis, el manejo y la explotación de dichos documentos son increíblemente altos. Los sistemas automáticos de exploración y análisis de redes de información contienen un desafío particularmente complejo; deben clasificar y perfilar los documentos para poder explotar, depurar y mejorar los sistemas de acceso a la información, así como los propios de análisis. A continuación se describe uno de los muchos escenarios en los cuales los algoritmos, y en particular ActiveRank, toman gran importancia dentro de los mismos sistemas de análisis de información, y que servirá como marco de referencia a lo largo del desarrollo de esta tesis.

Dentro de una red, se pueden presentar estructuras cíclicas que dificultan su análisis mediante sistemas automatizados, por lo que existe la necesidad de contar con métodos que sean capaces de detectar y manejar el comportamiento del analizador cuando se presentan dichas estructuras. Un ejemplo para entender claramente el efecto de las estructuras cíclicas en una red, y en el cual se centrará la implementación de este trabajo, es aquella formada por las páginas pornográficas en la red denominada World Wide Web; se trata de un conjunto de nodos altamente interconectados entre sí, y poco o nada interconectados hacia el exterior de su núcleo, es decir, podemos encontrar un vínculo para llegar a ellos, pero no un vínculo para salir de ellos, por lo que un sistema automático que entrara a dichas estructuras, no sería capaz de seguir escaneando otras secciones de la red, se vería atrapado si no fuera por su capacidad de detectar y manejar dicha situación.