



Universidad de Concepción - Chile

Facultad de Ingeniería

Departamento de Informática y Ciencias de la Computación

Una Estrategia de Navegación Asistida para Conjuntos de Documentos de Texto

Ramón Humberto Bilbao Salgado

Profesora Guía

María Andrea Rodríguez Tastets

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

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

Julio de 2007

Resumen

Cada día la cantidad de información disponible en la Web crece y, consecuentemente, se hace más difícil buscar la información deseada. Más aún, debido a la gran cantidad de documentos que pueden ser obtenidos de un proceso de recuperación de información, el recorrer la información recuperada puede ser un proceso tedioso y largo. Los avances en definir una estrategia de visualización de resultados, también llamado proceso de navegación de información, se han basado principalmente, en realizar agrupamientos (clusters) de resultados. Estos clusters, sin embargo, pueden aún así contener un gran número de documentos y no proponen un orden en el cual estos documentos pueden ser recorridos. El propósito de esta tesis es proponer un modelo que apoye la navegación de información, explorando técnicas que comparen documentos y modelen el espacio de información como una función de la diferencia entre documentos. El esquema de navegación finalmente propuesto utiliza la ganancia de información estimada de un documento aún no visitado con respecto a todos los documentos ya recuperados. Dicha ganancia de información es estimada como el número de nuevas unidades de información en documentos recuperados. De esta forma, la navegación es vista como una optimización respecto a la ganancia de información. El producto de la tesis es un sistema implementado que, usando un conjunto de documentos recuperados de la Web, sugiere una secuencia de recorrido de los documentos que optimiza la ganancia de información de un usuario. Los resultados obtenidos por la tesis indican que una técnica como ésta puede fácilmente complementar sistemas de visualización de resultados de búsqueda actuales y hacer que un sistema de recuperación alcance mayores niveles de eficiencia.