

Definición del Tema Amplio

Juan David Cruz Gómez

299618

Seminario de Investigación I

Maestría en Ingeniería – Ingeniería de Sistemas y Computación

5th September 2005

Minería de Datos para Extraer Conocimiento de Correos Electrónicos

Dado el crecimiento en la utilización del correo electrónico como medio ágil de comunicación, el volumen de información transportada a diario es muy alto, y esta información puede ser o correo no solicitado (SPAM) o correo autorizado no importante o correo autorizado de alta importancia. Este flujo permite la creación de redes de personas, o redes sociales, las cuales suelen tejerse en torno a un tema específico. Adicionalmente, los correos no son del todo excluyentes, y un mensaje puede tener uno o más temas, lo cuales pueden ser compartidos, de esta forma ya no se tiene una red social generada por los correos sino se puede tener también una red de temas, en general, crear un mapa conceptual utilizando la conexión existente entre los diferentes mensajes.

El conocimiento extraído de los mensajes, permite realizar diferentes tareas, entre ellas, filtrar el correo no deseado, aún más general, clasificar los mensajes de correo de acuerdo a temas o importancia, encontrar redes sociales y con estas, encontrar respuestas a la pregunta quién sabe que.

Para esto se propone la utilización de alguna técnica de minería de datos, que permita extraer la información necesaria para construir las redes, extraer y relacionar temas, clasificar el correo e identificar posibles personas que sepan de algún tema específico. La idea general es crear un grafo por tema, cuyos nodos sean personas y los arcos estén dados por la existencia de una relación, en este caso, en envío o la recepción de un correo.