

Generación automática inteligente de resúmenes de textos con técnicas de Soft Computing

Tesis doctoral en cotutela realizada por

Augusto Villa Monte

Directores:

Dra. Laura Lanzarini

Facultad de Informática - UNLP Argentina

Dr. José A. Olivas

Departamento de Tecnologías y Sistemas de Información

Universidad de Castilla-La Mancha Ciudad Real, España

RESUMEN

En la actualidad, Internet es el medio elegido para difundir información que, luego, es utilizada para resolver una amplia gama de problemas. Sin embargo, a medida que crece la cantidad de datos almacenados, su administración se vuelve aún más difícil y los usuarios comienzan a sufrir la llamada sobrecarga de información. Muchos son los sectores que, afectados por este fenómeno, no encuentran solución al problema. El uso, la disponibilidad y el desarrollo de tecnología en las últimas décadas han facilitado la recolección de información y también permitido la generación de grandes repositorios de datos. En los últimos años, los repositorios de documentos de texto, como lo es la web por ejemplo, han conseguido mayor atención. Dado el crecimiento exponencial en el volumen de información textual se volvió fundamental contar con herramientas automáticas que, a partir de la información original, diferencien lo esencial de lo que no lo es. No toda la información tiene el mismo nivel de relevancia. No sólo teniendo en cuenta su contenido sino también los intereses particulares. La obtención automática de resúmenes de texto constituye una solución a este problema especialmente en aquellas áreas de la ciencia como la medicina, en las cuales la investigación y divulgación de la información son fundamentales para su desarrollo. Esta tesis desarrolla dos estrategias diferentes para construir resúmenes automáticos de textos utilizando técnicas de Soft Computing. La primera de ellas es meramente reproductiva y la segunda más bien constructiva. Ambos casos implican fundamentalmente una jerarquización de la información y reducen el volumen del texto teniendo en cuenta el destinatario del resumen final. El énfasis estuvo puesto en la combinación de los enfoques. Por un lado, buscando resumir identificando el criterio que el usuario utiliza al seleccionar las partes relevantes de un 9 10 documento y, por otro, construyendo un grafo a partir de la captura de patrones textuales específicos sumamente útiles en la toma de decisiones.

Para la realización de los casos de estudio se utilizaron documentos de texto obtenidos de Internet de temática médica, área en la cual se realizó una aportación a través del diseño de una aplicación concreta.

Palabras clave: Resúmenes Automáticos, Procesamiento del Lenguaje Natural, Lógica Difusa, Cúmulos de Partículas, Causalidad, Inteligencia Artificial.