

# A framework and prototype for intelligent information retrieval

Vijay V. Raghavan  
Center for Advanced Computer Studies  
University of Southwestern Louisiana  
Lafayette, Louisiana 70504-4300

Ravi S. Sharma  
Department of Mathematics and Computing Sciences  
St. Francis Xavier University  
Antigonish, Nova Scotia B2G 1C0

Information Retrieval (IR) is the process of selectively disseminating relevant information stored among a variety of information items. A scheme for efficient (in terms of computer resources used) and effective (in terms of recall and precision) information retrieval is proposed. This draws from rough set theory<sup>1</sup> and an adaptive clustering strategy<sup>2</sup> in order to achieve some amount of self-organization in the database. The relevance judgments of weighted queries are used as the basis for influencing the classification of documents, thus enabling user input to direct clustering and allow the clustering sub-system to learn. An evaluation metric based on the theory of rough sets is used as the clustering criterion, which is enhanced in an iterative fashion. The clustering algorithm begins with an arbitrary

placement of the documents in the collection on a one-dimensional linear space and, as each query arrives, uses the relevance judgments to generate movement along the line so that "similar" documents are moved closer together and "dissimilar" ones moved further apart from each other. The precise definitions of what is meant by similar and dissimilar are also based on ideas from rough set theory. After a number queries are processed, the initial positions change as the system "learns" the users' profiles. The classification, however, is re-clustered only when this movement is deemed significant as defined by a variation of a common measure for determining the similarity between classifications. The general method of clustering places boundaries on gaps so as to optimise the clustering criterion. Efficient and effective retrieval is sought by using a learning scheme known as retrieval-by-example. The prototype experimental implementation of the scheme indicates remarkable potential, producing recall and precision results of between 40 to 80 per cent.

1. Pawlak, Zdzislaw. "On learning - A Rough Set Approach," Lecture Notes in Computer Science (No. 208, pp. 197-227). Skowron, R. (ed.), Springer-Verlag, Berlin (1986).

2. Yu, C.T., Wang, T.Y., & Chen, C.H. "Adaptive Document Clustering," Proceedings of the 8th Annual International Conference on Research and Development in Information Retrieval, pp. 197-203, ACM SIGIR, Montreal (1985).