International Journal of Innovative Computing, Information and Control
Volume **3**, Number **4**, August **2007**

ICIC International ©2007 ISSN 1349-4198
pp. **1031–1041**

EXTRACTIVE SUMMARIZATION USING A FUZZY-ROUGH-BASED RELEVANCE MEASURE

Hsun-Hui Huang, Yau-Hwang Kuo

Center for Researches of E-life Digital Technology
Department of Computer Science and Information Engineering
National Cheng Kung University
No.1, Ta-Hsueh Rd., Tainan, Taiwan
{ hhhuang; kuoyh }@ismp.csie.ncku.edu.tw

Horng-Chang Yang

Department of Computer Science and Information Engineering
National TaiTung University
No.684, Chunghua Rd., Sec.1, TaiTung, Taiwan
hcyang@nttu.edu.tw

Received October 2006; revised February 2007

Abstract. *In this paper, a novel method is proposed to extract key sentences of a document as its summary by estimating the relevance of sentences through the use of fuzzy-rough sets. This method uses senses rather than raw words to lessen the problem that sentences of the same or similar semantic meaning but written in synonyms are treated differently. Also included is semantic clustering, used to avoid selecting redundant key sentences. A prototype of this automatic text summarization scheme is constructed and an intrinsic method with criteria widely used in information-retrieval systems is employed for measuring the summary quality. The results of applying the prototype to a dataset with manually-generated summaries are shown.*
Keywords: Fuzzy-rough set, Word sense disambiguation, Semantic patterns retrieval, Key sentences extraction

1. **Introduction.** Automatic text summarization is a process which condenses a source document into a much shorter text while keeping its core content, and its importance to the development of search engines is obvious. The need for such a facility is made more acute by the huge amount and great variety of documents on the Internet. Though research on automatic text summarization can be dated back to 1950s [13][15] and the last ten years have seen much renewed attention given to this problem, the construction of an automatic summarization tool that produces coherent and cohesive summaries remains a challenge [6].

Approaches to automatic summarization can be divided into two fundamental categories - the knowledge-poor approach and the knowledge-rich approach [6] - which represent the endpoints of a continuum. The former is relatively shallow, while the latter is deep and complex. The knowledge-rich approach tries to analyze a text and turn it into a shorter one, using an array of tools such as the grammar, lexical databases and ontologies. The aim of this approach is, thus "summarization through abstracting". Apparently, the