The effects of analysing cohesion on document summarization

Rie Kubota Ando
Department of Computer Science, Cornell University, Ithaca, NY 14853-7501
kubotar@cs.cornell.edu

Branimir K. Boguraev, Roy J. Byrd, Mary S. Neff
IBM T.J. Watson Research Centre, P.O. Box 704, Yorktown Heights, NY 10598
bkb@watson.ibm.com,byrd@watson.ibm.com,neff@watson.ibm.com

Abstract This paper describes a framework for multi-document summarization which combines three premises: coherent themes can be identified reliably; highly representative themes, running across subsets of the document collection, can function as multi-document summary surrogates; and effective end-use of such themes should be facilitated by a visualization environment which clarifies the relationship between themes and documents. We present algorithms that formalize our framework, describe an implementation, and demonstrate a prototype system and interface.