Table based AHC Algorithm for Clustering Words
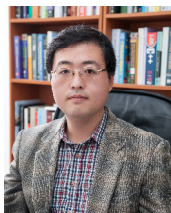
Taeho Jo

Department of Computer and Information Engineering, Inha University, 100 Inharo Namgu, Incheon, South Korea

tjo018@inha.ac.kr

Abstract—**This research proposes the table based AHC algorithm as the approach to the word clustering task. The results from encoding texts into tables were successful in the previous works on the text categorization and the text clustering, and if oppositely to the case of the text encoding, texts are assumed to be elements of each word, it becomes to be possible to encode words into tables. In this research, we map words into tables in the reverse direction to the text encoding, modify the AHC algorithm into the version where tables are given as its input, and apply the modified version of AHC to the word clustering task. In this research, we expect the better and more stable clustering performance, than the traditional version. Hence, this research provides the improved method of clustering words semantically.**

Keyword—**Word Clustering, Table based AHC, Table Similarity**

Taeho Jo (M'97–AM'12)   This author became a Member (M) of IEEE in 1997, and an Associate Member (AM) in 2012. He was born in 1970, South Korea. He received his Bachelor degree from Korea University in 1994, his Master degree from Pohang University of Science and Technology in 1997, and his PhD degree from University of Ottawa in 2006. His research area spans mainly over text mining, neural networks, machine learning, and information retrieval. He has the four year experience of working for industrial organizations and ten year experience of working for academic ones. So his research is characterized as the connection from fundamental researches for creating theories and to applied ones for developing products, by his experience of working for both sides.