

A Novel Approach to Compute Patterns Histories for Trend Analysis

Jing-Doo Wang

Department of Computer Science and Information Engineering,

Asia University

TaiChung, Taiwan

Email: jdwang@asia.edu.tw

September 11, 2010

Abstract

It is attractive to observe the variations of trends that are implicit in retrospective corpus over a long period of time. Pattern history could provide information analysts with valuable information and clues about the trend of some topics or the variation of writing/speaking styles. In this study one significant pattern is defined as a sequence of words that appears repeatedly and are enough to represent one event or topic; the history of one pattern is the frequency distribution of that pattern over time. To extract significant patterns from a large amounts of text, and meanwhile compute the corresponding patterns' histories, an external memory approach based on bucket-like suffixes sorting and push-pop stack operations is proposed. To highlight the scalability of this approach, there were 3,225,549 articles containing 677,728,269 words downloaded from the PubMed. Experimental results showed that patterns histories did reveal the variations of some events and gave hints for trend analysis.

Keywords: Text Mining; Trend Analysis; Pattern History; External Memory;