Detection Method to Prevent Reference Omissions Using Logs of Text Inserted by Copy-and-Paste Operations

Chikako Ishizawa*, Masahiro Toshima**, Yoichi Kageyama* and Makoto Nishida***

*Department of Mathematical Science and Electrical-Electronic-Computer Engineering,
Graduate School of Engineering Science,
Akita University, 1-1, Tegata Gakuen-machi, Akita-shi, Akita 010-8502, Japan
**Department of Computer Science and Engineering, Faculty of Engineering and Resource Science,
Akita University, 1-1, Tegata Gakuen-machi, Akita-shi, Akita 010-8502, Japan
***The Open University of Japan, 1-1, Tegata Gakuen-machi, Akita-shi, Akita 010-8502, Japan
E-mail:ishizawa@ie.akita-u.ac.jp

The objective of this study is to prevent human mistakes, specifically, omitting reference lists and reference numbers that should be appended to a copied text. When sentences are created using a personal computer (PC), it is possible to copy and use text from the Internet. However, a user will occasionally forget to include the copy-source information, e.g., the URL or webpage title, in the reference list or to add the reference number to the inserted text. Therefore, we propose a method that obtains the relevant logs from the PC, detects the text inserted by a copy-and-paste operation, and shows the copy-source information to the user. The proposed method consists of three processes. The first process monitors the clipboard state and creates logs with the copy-source information, e.g., the webpage title or URL. Thereafter, the text-detection process calculates the cosine similarity between the copy-source sentences and the user's text. Text with no reference information is detected based on the cosine-similarity value. Finally, the reference information for the detected text is displayed on the monitor. Various changes were made to the inserted text by using Microsoft Windows 7. Our experimental results showed that the proposed method can detect inserted text without reference information and display the necessary information in the reference list.

Keywords : Text detection, cosine similarity, log analysis