

## **CAText, Interactive Exploration of Chronological Text Corpora**

Nguyen-Khang PHAM\*

Can Tho University, Can Tho, Viet Nam, [pnkhang@cit.ctu.edu.vn](mailto:pnkhang@cit.ctu.edu.vn)

Annie MORIN

University of Rennes I, Rennes, France, [annie.morin@irisa.fr](mailto:annie.morin@irisa.fr)

In this paper, we present a novel graphical interactive tool CAText – Correspondence Analysis for Text. This tool is based on Correspondence Analysis (CA) coupled with visualization techniques, allows to visualize and to extract knowledge from textual corpora. Basically, CA produces a visual representation of the relationships between the documents and words in the same space. We extend its use in a recursive manner which enables corpus navigation and hierarchical topic detection. We also show how CAText can deal with chronological text corpora analysis although any type of temporal data can be applied. Enriched by meta-data filtering technique, CAText facilitates knowledge extraction process. All steps of an analysis process can be saved and transferred to other collaborators. Case studies on large corpora have shown the strength of CAText because of its intuitiveness, effectiveness, and robustness.

**Key Words:** Correspondence analysis, Visualization, Chronological text analysis