Trends in the Israeli State Comptroller - Text Analysis Using Data Analysis

Eran Raveh

Advisors: Dr. Ron Bekkerman and Dr. Hertzel Cohen - University of Haifa

The annual publication of the State Comptroller reports, become a media event, focuses on individual cases, not fully characterize the all issues included in the annual report. The larger question is, what is the scope of the issues that the state comptroller checks? In what does he focus? What changes in these trends over the years, when the head of State Comptroller office Replaced?

Qualitative studies of the State Comptroller's Office, pointed out different trends: Transition from traditional audit, focused proper operation of the government, to 'Advanced' audit of savings, efficiency, and effectiveness (Gates 1998); Shifts over the years, towards activism in State Comptroller reports (Hulata 2012).

This study have two goals: First - to characterize trends in the state comptroller over the years; Second - the development of research methodology that would allow characterization of trends, in large Text database of Hebrew documents, using quantitative research, based on Text Data Analysis .

This study performed in sequential steps: Step 1 - downloading data files from the Library of the State Comptroller, and structure arrangement of the text. Step 2 - construction of a dictionary of common terms (the "language"), listed in the State Comptroller (when database of Wikipedia is used as defining tool for relevant terms of the Hebrew language). Step 3 - Finding the optimal formula for performing classification of the State Comptroller's reports, based on the common terms in the documents (term frequency relative to document frequency). Step 4 - Data analysis and Data Visualization.

The contribution of this study is the development of a general methodology for quantitative and visual analysis, of all the documents included in big information Text database, in the Hebrew language that usually investigated in qualitative Studies.