Development of Visualization Search Service Model using Keyword Social Network for R&D Report

[1]YongJu Shin, [2]Hwan Gue Cho, [3]Ki Seok Choi, [4]ChulSu Lim, [5]Jiseong Son and [6]Kwang-Nam Choi

[1,3,4,5] NTIS Center, KISTI, Korea
[2] Dept. of Computer Science and Engineering, PUSAN National University, Korea
*6, Corresponding Author NTIS Center, KISTI, Korea

Abstract

As the amount of research data produced and used has increased rapidly due to advances in technology, there have been various attempts to handle it with Big Data and to provide useful results to researchers. Among these, large-scale documents such as R&D reports need a more efficient way to identify the contents and obtain relevant document information in a short time.

In order to quickly understand the contents of a document, visualizing the content of a R&D report in a single book will help to understand whether it is a desired R&D report. The model that visualizes and provides the contents of the R&D report proposed in this study is a document summary model that reflects the spatio-temporal characteristics and shows the dynamic change of words developed according to the content of the document. By applying this to the R&D report, the researcher can select a particular section in the document's time base to identify it in the form of a keyword graph.

In conclusion, this model can quickly identify the content of R&D reports that researchers can't understand at one time. Combining this model with the search service of existing research reports, it will help you quickly find the documents you need and the contents of R&D reports.

Previous Studies

Comparison of visualization tools

Tool	Layout	User Interaction	Mechanization of Document Analysis	Search within Document	Document Comparison	Document Stratification	Korean Language Support
Wordle	Cloud	Not available	Filters stop words	0	0	0	0
Voyant Tools	Cloud, Graph	Adds index and chart legend	Available	5	0	0	3
WordSift	Graph	Connects media search and graphs synonyms	Available	3	0	0	0
InfraNodus	Graph	Index, explores word relationship, connects with Google search	Available	4	0	0	0

※ Gives a score from 0 to 5 depending on the support of each function

Detailed Example of Document Visualization Tool

Wordle
- Visualizes document in a Word Cloud form
- Expresses frequency by size, poor delivery of information in an original text

<Inauguration speech of U.S. President Obama>

Voyant Tools
- Visualizes document in a Word Cloud form
- Frequency by each location

WordShift
- Arranges by the importance and emergence frequency of words
- Provides a graph connecting the synonyms of keywords and meaning-related words
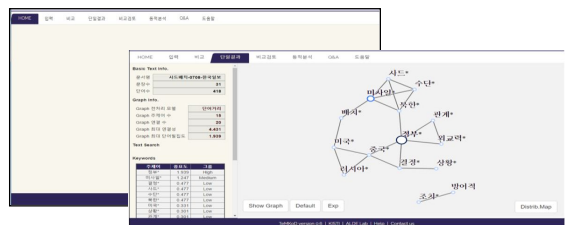

InfraNodu
- Calculates word connectivity using the simultaneous appearance of words
- Edge thickness varies depending on its weighted value

## Visualization Search Service Model for R&D Reports

User Interface

Main Functions

- Single Document Visualization Function
  - Extracts keywords and connectivity using word-distance model
  - Checks the single document graph created and keyword importance
  - Document reverse-search using keywords
  - Exports graph creation result
  - Keyword distribution visualization mode
- Dynamic analysis according to the document flow
  - Checks document flow according to phase-unit time and partial topic

Structure and Process
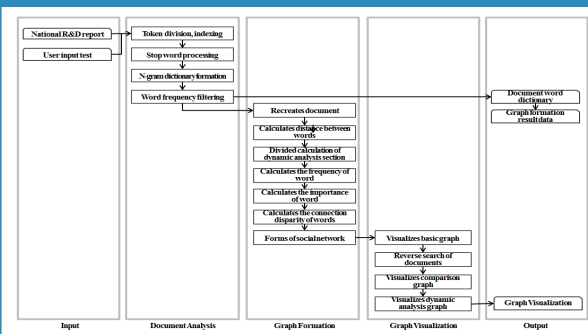
Result of Single Document Visualization

- Checks the basic information, word density and connectivity, and the importance of top N extracted keywords on a table.
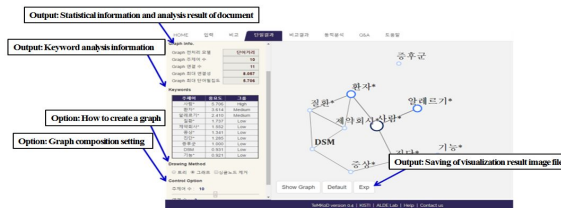
Result of Single Document Dynamic Analysis Visualization00

- Checks the basic information, word density and connectivity, and the importance of top N extracted keywords on a table.
- Checks the top 5 important sentences having the keywords by directly clicking the graphs.
- After moving to the 'Dynamic Analysis' tab, the document is divided to look at the graphs in phases to identify the detailed topic and the flow of document.
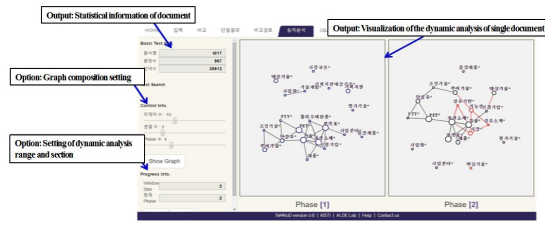
Conclusion

The R&D reports visualization search service model examined in this paper provides the analysis result of original R&D reports report text in real-time. Also, it is expected to be used as a high added-value service as it can enhance the understanding of document quickly, by grasping the changes of document according to time.

References

[1] Song Won-mun, Kim Young-jin, Kim Eun-joo, and Kim Myoung-won, "Automatic Document Summarization System using Dynamic Connection Graph", Journal of Information Scientists and Engineers: Software and Application Vol. 36 No. 1, pp.62-69, 2009.

[2] Kang Da-hyeon, Kim Jae-yoon, Park Kang-hee, Jeong Jae-won, Tak Hae-seong, and Cho Hwan-kyu, "Creation of Document Graphs using the Linear Overlapping Relation of Keywords", Collection of Published Papers in the Korean Institute of Information Scientists and Engineers, pp.592-594, 2013.