

Text Mining Applications in Discourse Analysis and Modern Trends (Bid Data Methods on Political Speeches)

Dagiimaa Balaanz /Associate professor, Ph.D/
MUNKHCHIMEG Otgonchuluun /Ph.D student, Linguistics/
At the University of Humanities, Ulaanbaatar, Mongolia

Abstract:- “*Superstructure*” or the macro structure is to define the super content model that applies to all structures from the smallest to the highest, or covers all levels from the base to the top. It includes all branched meanings beyond basic ideas and structures, or ideas derived from systems. The text represents the most important, core content based on the text. “Topic” refers to the general idea of the subtext that makes up any source text, or the “topic” that represents the core value of each complex. The concepts on “*mental model*” was created by Carsten Held in 2006. The mental model refers to the scope of understanding general knowledge in a wide range of areas: mind, knowledge, memory, experience, and information. The mental model is a key issue in language, behavior, and cognition. The mental model may be different for each individual’s perceptions, angles, and worldviews, and may be seen as an internal expression of the perception of external reality. Human cognition is a key strategy for the environment, events, decision-making, and problem-solving. In framework of the study, it tries to prove automatic text tools can export the same concepts as the humans do. The studies are processed in political speeches.

Political discourse analysis: In order to understand political discourse, it is necessary to study the following issues, including the expression of meanings of space, time, and nature. These concepts represent the realities of the world and of humanity. In other words, human beings use it to evaluate events, people, their behavior, places, and any phenomena that happen to them. This applies to the notion of self-esteem (Chilton, Schaffner, 1997), and space is an expression of materiality and metaphor. An approach to the issue of time or to a historical approach is also important. (ibid: 56) The concept of political discourse is related to time, such as yesterday, tomorrow, and before the war. (ibid) The concept of time, now or today, is the most important concept and determines other expressions depending on

how far and close they are from the center. (ibid: 58) In political discourse, this concept determines whether reforms are taking place and whether national past victories are present.

Keywords:- text mining, graph of word analysis, text automatic tools, Pearson’s correlation, Cronbach’s Alpha, macro structure, mental model.

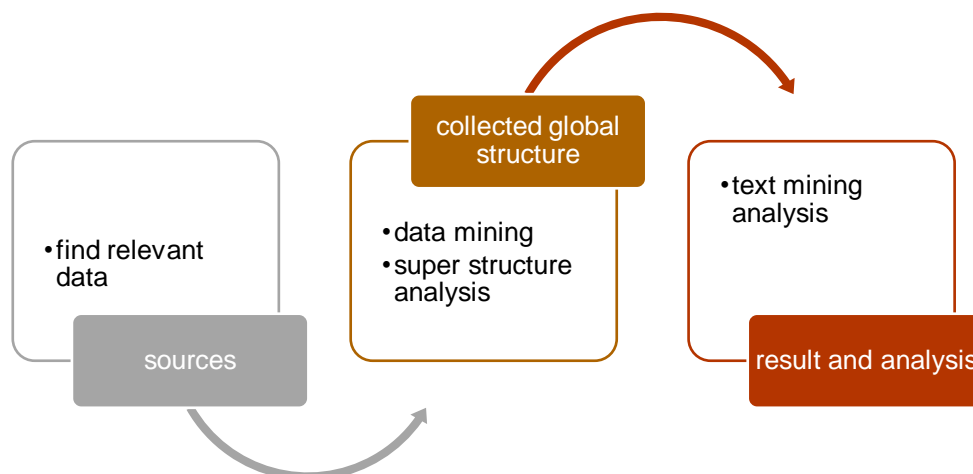
I. INTRODUCTION

A. Text mining analysis and its applications in discourse analysis and modern trends:

Modern “text mining” research methods have been used in research to organize and group documents. After that, the method became a ‘machine learning’ style of text-based research. Text mining is used to analyze text clusters, text macros, text sentiment analysis, and conversion of spoken text into numeric values. It is an analysis that summarizes any report, speech, text, and source material by means of text automation, and because of its importance in processing and analyzing information of interdisciplinary importance, the study used a computer-aided method in combination with linguistics. By collecting statistical data, developing a topical model, and mapping a language model, a structure that reflects linguistic features will be developed. Text mining calculations can be used to understand the relationship between a sentence, line, or complex, to predict, and to create categories and subgroups. How to do text mining Use text mining or text extraction techniques.

II. RESEARCH METHODOLOGY

It is the process of automatically generating the highest possible key values of valuable, weighted values, and limit values from any type of text to generate real-world concepts. By machine translating data into core information, text extraction automates the process of classifying text into content, meaning, topic, or topic. To illustrate this:



Graphic 1: Text mining process.

Text mining textual research methodology is developed in the following steps.

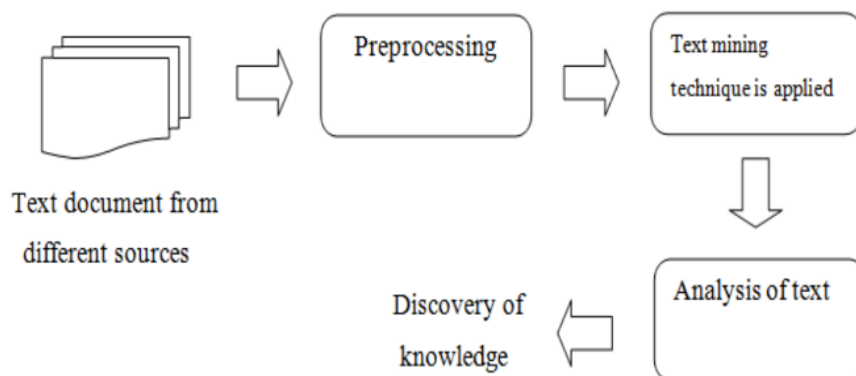


Fig. 1: Text mining processing scheme

III. DATA AND ANALYSIS

Text mining is a scheme of processing the text of our written language in the form of machine technology. In this chapter, the above reports are analyzed by “CA” or macrostructural component analysis. It also uses the “Graph of words, bag of words” algorithm to identify keywords and display targeted text content.

No.	Code	Name of speech	Meaning of speech
1	Oba S1 – /p1-	Barack Obama, Address at Fort Bonifacio delivered 29 April 2014, Manila, Philippines	Freedom, nations, commitment, peace, Americans vs Filipinos.

Table 1: Obama Speech 1-example:

In the speech on April 1, 2014, the concept is about the Philippine, and Manila, Obama highlighted the United States and the Philippines as allies, mentioning bilateral cooperation, military, counterterrorism, and counterterrorism. The main feature of his speech, the greeting of the speech, started very briefly with “HELLO

EVERYBODY”. First of all, Filinino proudly mentions the soldiers serving in the US military, noting that this day marks the 70th anniversary of the historic Battle of Leyte. At that time, the goal of thousands of warriors was peace and freedom.



Fig. 2: Obama Speech- key word for the “Word Cloud”

Complex concept	Correlation	Frequency of words
armed forces	97	3
filipino veterans	96	2
american troops	96	2
philippine navy seal	95	2
filipino resistance fighters	93	1
american troops loading	92	1
filipino soldiers unloading	92	1
american cargo aircraft	90	1
states marines	90	1
carrying captain	89	1
defining moments	88	1
mutual defense treaty	88	1
vice president	87	1
nations stand	84	1
global relief effort	84	1
disaster zone	82	1
death marches	82	1
colonel mike	80	1
filipino helicopters	79	1
filipino friends	78	1

Table 2

Cronbach’s Alpha analysis for the data above.

Reliability Statistics	
Cronbach’s Alpha	Key concepts above
0.752	5

Source: SPSS analysis by the

Table 3: Reliability analysis of the “Barack Obama” survey report

The value of this analysis is 0.752, which indicates that the reliability is “good”.

Attitude expressing sentiment	Percentage
Opinion	71%
Fact	29%

Fig. 3: “Barack Obama” speech sentiment analysis



Fig. 4: "Barack Obama" speech sentiment analysis by graphic result

Sentiment analysis: Sentiment analysis is designed to determine how a politician expresses his or her emotions in a speech and whether the report is based on the author's basic knowledge or factual evidence. Of these, 71 percent of the presentations were statements of opinion, 29 percent were facts, and the balance of "censorship" was positive. There's a connection between our proud veterans from World War II and our men and women serving today -- bound across the generations by the spirit of our alliance, Filipinos and Americans standing together, shoulder-to-shoulder, balikatan and he reiterated his satisfaction with the progress made by the joint efforts.

Oba S3: Barack Obama - Announces Candidacy 2007

When downloading the data from the text automation database on relevant tools, you will find the following types of related words with possible variants of nouns and symbols. This is defined as the target "limit values".

You know, we all made this journey for a reason. It's humbling to see a crowd like this, but in my heart I know you didn't just come here for me. You came here because you believe in what this country can be. In the **face of** war, you believe there can be peace. In the **face of** despair, you believe there can be hope. In the **face of** a politics that shut you out, that's told you to settle, that's divided us for too long, you believe that we can be one people, reaching for what's possible, building that more perfect union.

IV. SUMMARY

'In the face of a politics that shut you out, that's told

you to settle, that's divided us for too long, you believe union'.that we can be one people, reaching

for what's possible, building that more perfect

Topic -1 lexical word in use:

Topic proposition: 'In the face of a politics that shut you out, that's told

you to settle, that's divided us for too long, you believe union'.that we can be one people, reaching

for what's possible, building that more perfect ": to create text automation

The superstructure of discourse can be seen as follows. The superstructure of discourse is the key word that conveys global meaning.

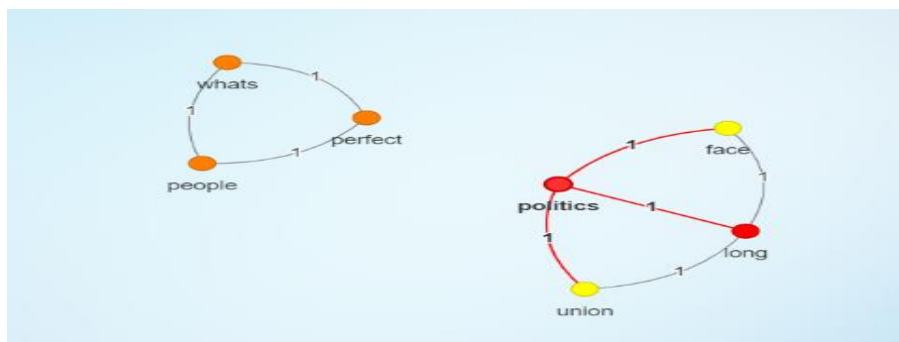


Fig. 5: "Obama" speech keyword or GoW

- FACE OF POLITICS /will be peace, despair, hope, perfect, union/
- PEOPLE /can be perfect union/

Elbegdorj(4th president of Mongolia) Speech – example 2**United Nations General Assembly****Sixty-fourth session****8th plenary meeting Friday, 25 September 2009, 3****p.m. New York****Highlighted persons**

Name	Correlation	Frequency
Barack Obama	27%	1

Table 2: names with correlation by frequency in data

Reliability Statistics

Cronbach's Alpha	Газар орны нэр
.921	7

Table 4: Cronbach alpha

The value of this analysis is 0.921, which indicates that it is “very good” in terms of reliability: Topic by directed key words: Allow me to **share** briefly our **views** on **issues** we deem **important** as we collectively seek to identify effective **responses** to the **global crises**. First, my **delegation** believes that the **multiple nature** of the **crises**

has to be taken into **account** in order to find an **adequate response** at the **global level**. This in itself is a **daunting task**, requiring of us the **courage** to rise beyond mere **national** or **group** interests in order to **survive** collectively in our one — **global** — **human village**.



Fig. 6: “Obama” speech keyword or GoW

Topic value – concepts: Identify the words with the highest rank in the data:

Topics	Weighted value
Democracy	99.644
Nuclear Policy	92.3827
Nuclear Weapons	92.2958
Human Rights	91.1578

Table 4: topic words and value

Reliability Statistics

Cronbach's Alpha	Topics
.691	4

Table 5: Cronbach alpha.

The value of this analysis is 0.691, which can be considered as “reasonable” reliability. The following proposition, which is the core of the content of the above idea, can be named: Despite this grim situation we are

nonetheless encouraged by a broadly shared recognition that the vulnerable countries, including landlocked developing countries (LLDCs) ought to be assisted to withstand the harsh impact of the crises. In this regard, we look forward to

the Group of 20 meeting in Pittsburgh to substantially increase support and assistance to vulnerable and low-income countries. When these key words are constructed by text automation, which defines the macro structure of the meaning for the speech, the superstructure of the discourse

can be seen as follows. The superstructure of discourse is the key word that conveys global meaning.

Summary of the CONCEPT: Here are the high-ranking keywords that define the super value of a section:

- **CONCEPT - propositional key words in word cloud:**



Fig. 7: “Elbegdorj Ts”speech key word in Word Cloud.

Allow me to **share** briefly our **views** on **issues** we deem **important** as we collectively seek to identify effective **responses** to the **global crises**. First, my **delegation** believes that the **multiple nature** of the **crises** has to be taken into **account** in order to find an **adequate response** at

the **global level**. This in itself is a **daunting task**, requiring of us the **courage** to rise beyond mere **national** or **group interests** in order to **survive** collectively in our one — **global** — **human village**.

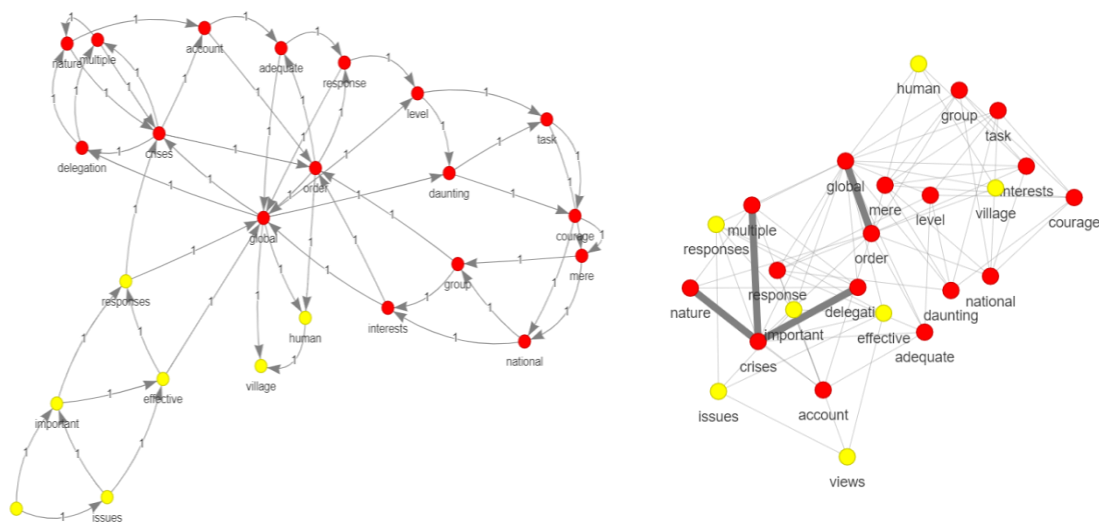


Fig. 8: “Elbegdorj Ts” speech key word in Word Cloud.

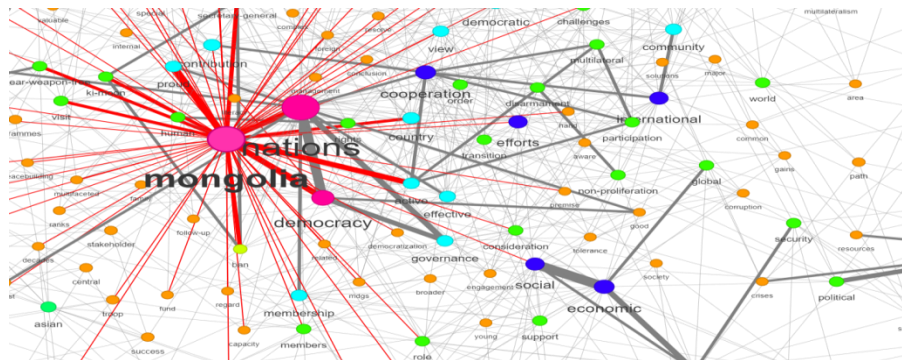


Fig. 9: “Elbegdorj Ts” speech key word in Word Cloud

V. SUMMARY

This single-topic research paper is being developed and summarized. In the field of linguistics, the use of language, vocabulary, speech, and semantic structure are important for discourse research. Thus, in combination with the emerging text mining methods of discourse research, the study was conducted using text automation to improve discourse analysis on the example of political discourse and speeches. In the course of the research, data and information are collected using digital tools to develop analysis at the macro and micro levels of discourse, and the researcher's qualitative analysis, as well as statistical and quantitative analysis, are developed and presented.

- In the development of discourse analysis, it is emphasized that the use of modern textual text mining analysis methods is effective in recognizing and understanding the general idea of speech and presentation at the macro level using “topics” and “keywords”. He noted that creating superstructures or keywords to create a “topical model” is more important for understanding the main idea and concept in a shorter period of time than reading, listening to, or viewing the source material. By creating a macro model of speech and using automation methods in discourse research, the spread of keywords was analyzed by a scatter plotter to understand the main ideas, propositions, and sub-meanings of the text.
- Native language speakers and foreign language presentations show differences in sentence style, speech subtopics, and propositional levels. The hypothesis that it is directly related to education, ideology and basic knowledge has been confirmed.
- The use of automated methods of discourse analysis in discourse analysis can be used to identify super-structures using modern 'text mining' methods to understand the global meaning of political discourse. There will be. By reading and understanding speech and text; Determining the super-structure using the “text mining” method determined that there was no difference in content.

REFERENCE

- [1.] Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, Miller, C. Sotillo, H. S. Thompson and R. Weinert, *The HCRC Map Task Corpus. Language and Speech*, vol. 34, 1991, pp. 351-366.
- [2.] C. Graesser, K. VanLehn, C. Rose, P. Jordan, D. Harter, *Intelligent tutoring systems with conversational dialogue*, *AI Magazine*, 2001, 22, 39-51.
- [3.] C. Graesser, P. Chipman, B. C. Haynes, and A. Olney, *AutoTutor: An intelligent tutoring system with mixed-initiative dialogue*, *IEEE Transactions in Education*, pp. 612-618, 2005.
- [4.] C. Mengel, L. Dybkjaer, J. M. Garrido, U. Heid, M. Klein, V. Pirrelli, M. Poesio, S. Quazza, A. Schiffrin, and Soria, *MATE Dialogue Annotation Guidelines - MATE Deliverable D2.1*, Technical Report, 2000.
- [5.] Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, M. Meteer, and C. V. EssDykema. 2000. *Dialogue Act Modeling for Automatic Tagging and Recognition of onversational Speech*, *Computational Linguistics* 26:3, 339-371
- [6.] G. Salton and C. Buckley, “Term-Weighting Approaches in Automatic Text Retrieval,” *Information Processing and Management: An Int’l J.*, vol. 24, no. 5, pp. 513-523, 1988.
- [7.] H. Ahonen, O. Heinonen, M. Klemettinen, and A.I. Verkamo, “Applying Data Mining Techniques for Descriptive Phrase Extraction in Digital Document Collections,” *Proc. IEEE Int’l Forum on Research and Technology Advances in Digital Libraries (ADL ’98)*, pp. 2-11, 1998.
- [8.] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins, “Text Classification Using String Kernels,” *J. Machine Learning Research*, vol. 2, pp. 419- 444, 2002.
- [9.] H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition ed. San Francisco: Morgan Kaufmann, 2005.
- [10.] J. Helbig, B. Schindler, *Speech-Controlled Human Machine Interaction (Sprachgesteuerte Mensch-Maschine-Interaktion)*, *Information Technology*, Vol: 46, Issue: 6-2004, Pps: 291 – 298, Oldenbourg Wissenschaftsverlag GmbH.
- [11.] J. Horecký, *The Content and Form of Illocutionary Acts*, *SKASE Journal of Theoretical Linguistics* [online]. 2007, vol. 4, no. 1
- [12.] J. L. Austin, *How to Do Things with Words*, Oxford: Oxford University Press, 1962.
- [13.] K. VanLehn, A. C. Graesser, G. T. Jackson, P. Jordan, A. Olney, and C. P. Rose, *When are tutorial dialogues more effective than reading*, *Cognitive Science* (in press).
- [14.] P. Taylor, S. King, S. Isard and H. Wright, *Intonation and Dialogue Context as Constraints for Speech Recognition*, *Language and Speech*, vol. 41, 1998, pp. 493-512
- [15.] R. Sharma, M. Yeasin, N. Krahnstöver, I. Rauschert, G. Cai, I. Brewer, A. MacEachren, K. Sengupta, *Speech-Gesture Driven Multimodal Interfaces for Crisis Management*, *Proceedings of IEEE (special issue on Multimodal Human-Computer Interface)*, Vol.91, No.9, pp. 1327-135, 2003.
- [16.] S. Joty, G. Carenini, R. T. Ng and G. Murray, “Discourse Processing and Its Applications in Text Mining,” *2018 IEEE International Conference on Data Mining (ICDM)*, 2018, pp. 7-7, doi: 10.1109/ICDM.2018.00014.
- [17.] S. Shehata, F. Karray, and M. Kamel, “A Concept-Based Model for Enhancing Text Categorization,” *Proc. 13th Int’l Conf. Knowledge Discovery and Data Mining (KDD ’07)*, pp. 629-637, 2007
- [18.] S. Shehata, F. Karray, and M. Kamel, “Enhancing Text Clustering Using Concept-Based Mining Model,” *Proc. IEEE Sixth Int’l Conf. Data Mining (ICDM ’06)*, pp. 1043-1048, 2006.
- [19.] S.-T. Wu, Y. Li, and Y. Xu, “Deploying Approaches for Pattern Refinement in Text Mining,” *Proc. IEEE Sixth Int’l Conf. Data Mining (ICDM ’06)*, pp. 1157-1161, 2006.

- [20.] S.-T. Wu, Y. Li, Y. Xu, B. Pham, and P. Chen, "Automatic Pattern- Taxonomy Extraction for Web Mining," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI '04), pp. 242-248, 2004.
- [21.] T. Hofmann, EECS Department, Computer Science Div., Univ.of California, Berkeley & International Computer Science Institute, Berkley, CA, Probabilistic Latent Semantic Analysis– Uncertainty in Artificial Intelligence, UAI'99, Stockholm
- [22.] W. Lam, M.E. Ruiz, and P. Srinivasan, "Automatic Text Categorization and Its Application to Text Retrieval," IEEE Trans. Knowledge and Data Eng., vol. 11, no. 6, pp. 865-879, Nov./Dec. 1999.