Generating mind map from an article using machine learning

M F Nurrokhim, L S Riza and Rasim*
Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

*Corresponding author’s email: rasim@upi.edu

Abstract. Sequence teaching materials is very important in learning because it will improve the effectiveness of learning. But not all teachers can make the sequences in detail and accurate. The purpose of this research generates sequences of teaching materials automatically by using the concept map. The approach used is the Information Retrieval Approach and the use of the CoreNLP library. Concept map is generated from articles or other teaching materials. Testing is done by comparing the concept map of automatic generation with concept map made by experts. The results showed 52.5% accuracy.

1. Introduction
The preparation of teaching materials is an important process in learning. This preparation takes a long time and big energy. The use of mind map for the determination of teaching materials can be very helpful because the mind map compiled the material with a structured. This makes it easier for the teacher to determine the order of the material to be taught. The applications that make the mind map is Xmind, FreeMind, Edraw MindMap, and MindMeister. But this application is manual, because the user manually determines the mind map element. Automated mind map generation has been done on Indonesian text using NLP. Abdeen et al., 2009 generate the mind map from a text with M2Gen. Mind map in this research has the elements can be: title, subtitle, and topic sentence. Determining the topic sentence manually of each paragraph certainly takes a lot of time. Automatic generation of topic sentence based on syntax has been done by He and Peng, 2013 while Asawaroengchhai et al., 2018 generated the topic sentences with Topic N-grams model (TNG) for Thai language.

Research on the mind map does generate the mind map of the text. This research generated the mind map of an instructional source in the form of articles that have the structure: title, subhead, and paragraph. Each paragraph in the article specifies the topic of the sentence which is then integrated into the mind map. The resulting mind map is an outline of mind map with structure: after getting topic sentence from each paragraph will generate an outline of mind map title, sub-title, sub-sub-title, and topic sentence. Generating a mind map using the approach Information Retrieval Approach and NLP.

This research is organized into 4 sections, namely: introduction, research method, research results and discussion, and conclusion. Introduction contains background issues and research objectives. The research method contains the research steps. The results of research and discussion contains the results
of research in the form of generating mind map for learning. The conclusion contains the recapitulation of the research discussion.

2. Methods

Based on figure 1, it can be seen that this research is done with 6 stages namely:

1. Literature Study: is the first stage of the researcher to search, study and understand the theories and methods used such as: learning, automatic generating mind map, Natural Language Processing, Core NLP, Information Retrieval Approach.

2. Data Collection: The data used in this study consists of two types, namely:
   a. Article: article consists of several paragraphs, data in given to the expert for validation.
   b. Validation Results: articles that have been marked topic sentence in each paragraph.

3. Model design: At this stage there are several things that are developed that is Information Retrieval Approach where in it there is a pre-processing text process, Core NLP, Part of Speech Tagger. At the process of text processing technique used is segmentation, tokenization. Data is processed to separate sentences into words per word text pre-processing, then done core NLP and part of speech tagger to determine the word type per word in the sentence. There are three types of words specified, among others, Noun, Verb, and Adjective. The next process is the feature extraction in which there is a process: frequency, sentence position value, and similarity with title and sub-title. After that, sentence scoring is done to sum up the value obtained from the feature extraction. Then the last stage is the sentence ranking and summary extraction after the calculation is done it will be selected the greatest value to be the topic sentence of the paragraph.

4. Application Development: This application development uses linear sequence process model with stages (requirement definition (analysis), system and software design, implementation and unit testing, integration and system testing and operating and maintenance stage support) [6].

5. Experiments: experiments conducted with the steps: a) determine the criteria of the article and b) the process of determining the topic of sentences by the experts and applications.

6. Research Results: Experimental results are articles with topic sentence every paragraph.

3. Results and Discussion

Teaching materials are learning materials for students. Teaching materials can be made from books or articles with attention to learning objectives. In this study, learning resources used are articles. Articles used in this study were taken from https://greatergood.berkeley.edu/education and https://markmanson.net/best-articles totaling 15 Articles. While the expert to validate is 2 people, namely: Mrs Tati Sugianti, S. Pd while for the second expert named Mr. Ahmad Fatoni. The automatic generating mind map process follows the path as shown in Figure 3.

1. Text Preprocessing. In this process, the document will be parsed with the following order: (1) Title; (2) Subheadings; (3) Sub-headings; and (4) Paragraphs.
2. **Tokenization** using Core NLP. In this process the paragraphs that have been diacase folding will be split into sentence then into a word that is loaded in a 3-dimensional array.

3. **Feature Extraction**: in this process there are three stages namely: (1) **Frequency**; (2) **Sentence Position Value**; (3) **Similarity with the Title**.

4. **Sentence Scoring**: equations for finding sentence scoring results.

\[
SS = \text{Freq} + \text{SP} + (\sum \text{ST})
\]  

Note:  

\begin{align*}
SS & = \text{Sentence Scoring} \\
\text{Freq} & = \text{Frequency} \\
\text{SP} & = \text{Sentence Position Value} \\
\text{ST} & = \text{Similarity with the title and sub-title}
\end{align*}

5. **Sentence Ranking**: sorting sentences by the greatest value.

6. Summarization of the processed paragraph: The result of Summarization will be combined with the previous outline so that Final Outline is obtained.

Based on the test result in one article, obtained topic sentence generated by the system presented in Figure 2.

![Figure 2. Generating sentence topics.](image)

Figure 3. shows the outline of the mind map generated by the system. In the figure can be seen that the outline has a structure, namely: the title of the article, sub-title, sub-subtitle and so on, and \( n \) the topic of the sentence. The result of the mind map of the expert and system is then performed analysis to determine the accuracy of the system.
RESULTS of OUTLINE MIND MAP

THE WAYS TO CULTIVATE HOPE

1. become mindfully aware of what’s going on inside

Paragraf 1
in order to change our beliefs about ourselves, we have to first
know what they are

Paragraf 2
the practice of mindfulness can help us observe that our bodies or
emotions are telling us something is not quite right, which then
allows us to describe what we are experiencing

2. be gentle with yourself and change your narrative

Paragraf 1
sometimes, though, it can be difficult to develop the “acting with
awareness” phase of mindfulness

Paragraf 2
but it’s not enough to soothe yourself with kindness

After getting the experimental results data, then the next step is to calculate the average of the correct
and wrong topic sentence generated system. The results will be presented in a table with the aim of
making it easier to analyse as illustrated in Table 1.

Table 1. Analysis of experimental results.

| ARTICLE | TOPIC NUMBER | SCORE | EXPERT 1 | EXPERT 2 |
|---------|--------------|-------|----------|----------|
|         |              |       | TRUE     | FALSE    |
|         |              |       | TRUE     | FALSE    |
| 1       | 15           | 10    | 5        | 10       |
| 2       | 13           | 9     | 4        | 9        |
| 3       | 12           | 4     | 8        | 5        |
| 4       | 9            | 6     | 3        | 3        |
| 5       | 11           | 9     | 2        | 6        |
| 6       | 7            | 4     | 3        | 4        |
| 7       | 9            | 5     | 4        | 4        |
| 8       | 8            | 4     | 4        | 3        |
| 9       | 10           | 4     | 6        | 5        |
| 10      | 6            | 4     | 2        | 3        |
| 11      | 10           | 4     | 6        | 3        |
| 12      | 7            | 3     | 4        | 5        |
| 13      | 7            | 5     | 2        | 6        |
| 14      | 18           | 7     | 9        | 7        |
| 15      | 22           | 11    | 11       | 10       |

AVERAGE (%) 55.4% 44.6% 51.7% 48.3%

Figure 3. Outline mind map according to the following articles:
https://greatergood.berkeley.edu/education and https://markmanson.net/best-
articles.
Table 1 shows that there are 2 experts who determine the accuracy of the system. The true and wrong values of each expert are calculated from the number of paragraphs in the article. The first expert had an average true value of 0.554 and an average error value of 0.446. While the second expert has an average true value of 0.517 and the average error value of 0.483. From this it can be analyzed that the percentage of the average value of both experts is true value of 53.55% and the error is 46.45%. It means that even though the system needs to be improved, it has been in the correct track to be used for determining main topics based on the full articles.

4. Conclusion
Automatic generating mind map produces teaching materials with the following structure: title, subtitle and topic of sentence every paragraph. The accuracy of the system is derived from the true average value that the expert produces on the results of the system. The value obtained is 55.4% for the first expert and 51.7% for the second expert with an average of 53.55%. this indicates the accuracy of the system is medium. In the future, we can combine this proposed system with other applications in learning media (such as [7-11]) and educational evaluations (e.g., [12]).

5. References
[1] Saelan A and Purwarianti A 2013 Generating mind map from Indonesian text using natural language processing tools Procedia Technology 11 1163.
[2] Abdeen M, El-Sahan R, Ismaeil A, El-Harouny S, Shalaby M and Yagoub M C E 2009 Direct automatic generation of mind maps from text with M 2 Gen Proc 2009 IEEE Toronto International Conference In Science and Technology for Humanity (TIC-STH), p. 95
[3] Arnaudet M L and Barrett M E 1981 Paragraph: a guide for students of English as a second language (Englewood Cliffs, N.J.: Prentice-Hall)
[4] Paul W 2002 Sentence-internal topics in Mandarin Chinese: The case of object preposing Language and Linguistics 3 4 695
[5] Asawaroengchai C, Chaisangmongkon W and Laowattana D 2018 Probabilistic learning models for topic extraction in Thai language. Proc 2018 5th International Conference on Business and Industrial Research (ICBIR), p. 35
[6] Sommerville I 2011 Software Engineering (Addison-Wesley)
[7] Widiaty I, Riza L S, Ana A, Abdullah A G, Mubaroq S R and Abdullah M 2018 Designing a content setting application on youtube in learning batik in a vocational school. Proc IOP Conference Series: Materials Science and Engineering 434
[8] Widiaty I, Yustiawan I, Wibisono Y, Abdullah A G, Abdullah C U, Riza LS 2018 Implementation of Markerless Augmented Reality Method to Visualise Philosophy of Batik Based on Android Pertanika Journal of Science & Technology 26 3.
[9] Riza L S, Firdaus D S, Junaeti E, Hidayat T, Abdullah A G, Nandiyanto A B, Abdullah C U 2018 A concept and implementation of instructional interactive multimedia for deaf students based on inquiry-based learning model Journal of Engineering Science and Technology 13 7 2016
[10] Riza L S, Putro B L 2018 Development of cyberglog-based intelligent tutorial system to improve students learning ability algorithm Proc Journal of Physics: Conference Series p. 012157
[11] Fauzia N, Rohendi D, Riza L S 2016 Implementation of the Cellular Automata Algorithm for developing an educational game. Proc 2nd International Conference on Science in Information Technology (ICSITEch) p. 169
[12] Farasyi G, Setiawan W, Fahsi M, Riza L S 2018 The Association Rule Method for Mapping and Recommendation System on Students’ Difficulties Transylvanian Review 1 1.