

The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)

Generating Mind Map from Indonesian Text using Natural Language Processing Tools

Athia Saelan*, Ayu Purwarianti

Bandung Institute of Technology, Jl. Ganesha no. 10, Bandung 40132, Indonesia

Abstract

People usually make mind map by drawing each object and its relation with other object from scratch. This research aims to make the process easier by generating mind map from text (here is Indonesian text) and providing mind map editor to manipulate the object and relation set. To build such tool, we employ available Indonesian NLP (Natural Language Processing) tools. There are three components needed: semantic net generator, mind map visualization and interaction handler. In the semantic net generator, the resulted first order logic (FOL) resulted by the semantic analyzer is changed into semantic net which is represented by list of objects and list of relations. The resulted semantic net is then visualized by using combination method of radial and layering drawing. The interaction is available for editing the object and the relation. The tool was then evaluated by 2 experiment set: testing the semantic net generation and testing the resulted visualization. The semantic net generation was evaluated by using the valid input text, while the visualization was evaluated by user acceptance test. As the result, although the semantic net generation (from FOL) is a correct one, but the whole semantic analyzer for Indonesian text still has a low accuracy especially for complex sentence. As for the user acceptance test, the automatic generation still gives unimportant object which should be corrected by the interaction.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](#).

Selection and peer-review under responsibility of the Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia.

Keywords : mind map; Indonesian; semantic; visualization; first order logic

* Corresponding author.

E-mail address: athiasaelan@yahoo.co.id

1. Introduction

Nowadays, information is easy to be collected, especially by the internet technology development. People can gain their needed information just in second. Moreover, now books are available in digital form.

Unfortunately, most information is available in text which is not easy to be caught by human mind. On the other hand, picture representing information is easy to be caught by human mind. Mind map is a concept representing this opinion which is effective to make human mind understand and remember information. By picture and color combination, mind map can represent fact and thought which is compatible with how the brain works[1].

Even though people understand mind map's benefit, still not many people make use of mind map since it is troublesome and wasting time, especially when the text size is large.

2. Related Works

Descriptions below are about related mind map generator for English, namely M²Gen[2] and Actor-based Mind-Maps Assembler[3].

2.1. M²Gen[2]

There are three main components in M²Gen: natural language processing, mind map conversion and mind map view manager. The architecture is shown in Fig. 1. In the morphological analysis, the process is to analyze each word into its lemma and affix along with its POS tag. In the parsing, the grammar is written in CFG form while the algorithm is the top down chart parsing. Not all parse tree are used in the semantic analyzer, therefore the syntax analysis filter the parse tree result into the needed information. Next process is the semantic analyzer which includes discourse analysis, word sense disambiguation and text meaning representation. The resulted semantic model is then converted into mind map figure. The example of generated mind map figure is shown in Fig. 2

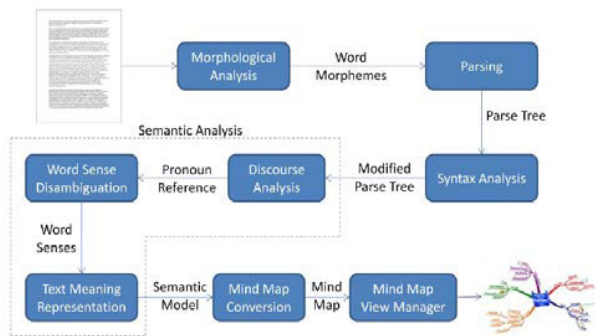


Fig. 1. Process Architecture in M2Gen [2]

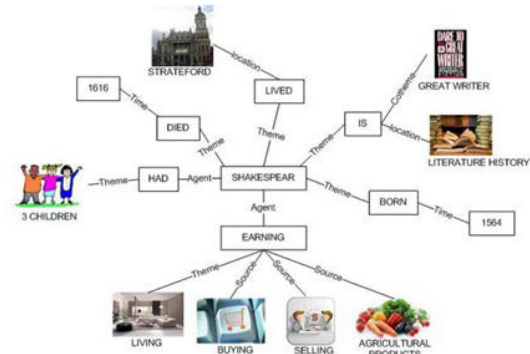


Fig 2. Mind Map Generated by M2Gen [2]

2.2. Actor-based Mind-Maps Assembler[3]

This research assumes that the main concept of a text is the actor. Sentence subject and object become the concept, while the sentence predicate is the concept relation. The adjective becomes the sub-concept.

The whole process is shown in Fig. 3. preprocess is a component to get the syntactical and semantic information from an input text. The result is then processed by pronoun resolution to search the sense of a pronoun. Next process is to take the subject-verb-object from each sentence along with relation between sentences. The result is represented in semantic network. Last, the co-reference resolution component joins the same concept among sentences. The mind map example is shown in Fig. 4.

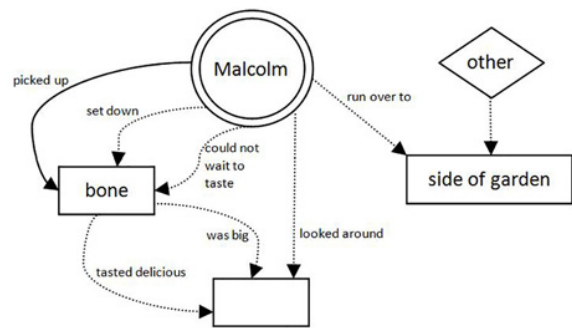
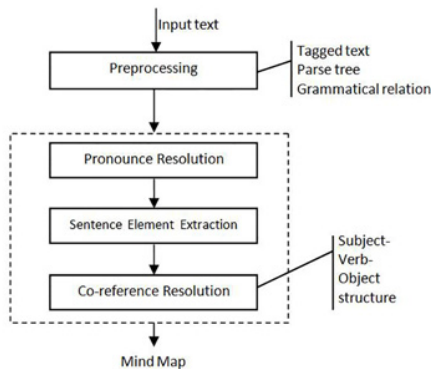


Fig. 3. Process Architecture in Actor-based Mind Maps Assembler[3]

Fig. 4. Mind Map resulted by Actor-based Mind Maps Assembler[3]

3. Indonesian Mind Map Generator

3.1. The Semantic Representation

Here, we tried to analyze the previous researched mind map representation [2][3] into Indonesian text as below: “Kartini lahir di Jepara. Jepara berada di Jawa Tengah. Kartini lahir pada tanggal 21 April 1879. Beliau adalah tokoh. Kartini mendirikan sekolah perempuan pada tahun 1913. Sekolah itu bernama Sekolah Kartini. Kartini menulis surat-surat.”

In M2Gen, all words become the node which in Actor-based Mind-Maps Assembler, the noun becomes node and verb becomes relation. The mind map representation in M2Gen is shown in Fig. 5 and the mind map representation in Actor-based Mind-Maps Assembler is shown in Fig. 6.

Basically, both methods are similar, but the first method is more flexible because it is easier to represent a sentence having no object or sentence with a lot of objects. For example, for sentences: “Kartini lahir di Jepara” and “Kartini lahir pada tanggal 21 April 1879”, the second method will gives two branches while the first method will results one branch. We selected the semantic text representation proposed by the first method which can be seen as a type of semantic network. As for the internal data structure for this representation, we chose to use a list of objects and a list of object relations. For example, the objects are “Kartini”, “lahir”, “Jepara”, and the relations are “Kartini – lahir” and “lahir – Jepara”.

3.2. Alternative Method to Generate the Semantic Representation

The important consideration here is that the process appropriateness with the needed semantic representation and the availability of Indonesian Natural Language Processing (NLP) tools. In the M2Gen, the method is to use all important NLP tools, ranging from the morphologically analysis to semantic analyzer. In the Actor-based Mind-Maps Assembler, the employed method is the syntactical analysis. Another alternative is the name entity extraction and relation extraction. By this, there are 3 alternative methods: using semantic analysis, syntactical analysis and named entity-relation extraction.



Fig. 5. Mind Map for the Indonesian Sentence such as in M2Gen

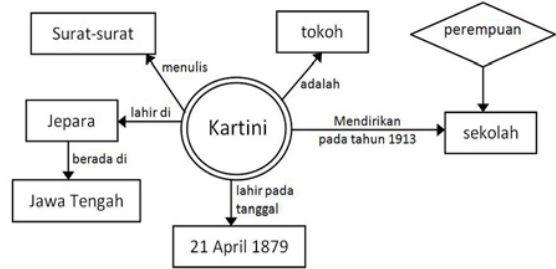


Fig. 6. Mind Map for the Indonesian Sentence such as in Actor-based Mind Maps Assembler

3.2.1. Syntactical Analysis

There are two common syntactical analysis or parsing methods: constituency based and dependency based. Since the constituency based gives result of phrase list and the dependency based gives result of word relation list, we concluded that the dependency based is more appropriate for the mind mapping generator with easier process to transform the dependency parse tree into the list of objects and list of object relations. To evaluate it further, we analyzed the parse tree for each sample sentence mentioned before. Unfortunately, the available Indonesian dependency parser is not equipped with its reference resolution which makes it difficult to see the sentence relation.

3.2.2. Semantic Analysis

In M²Gen, the semantic analysis gives result of semantic network. For Indonesian language, the available semantic analysis gives result of FOL[4]. This semantic analysis already has its anaphora resolution. The FOL resulted for the sample Indonesian sentence is shown in Table 1. The FOL can then be transformed into the semantic networks by defining mapping rules between terms in the FOL and object/relation in the semantic networks.

Table 1. Examples of FOL for Indonesian Sentence.

Sentence	FOL
Kartini lahir di Jepara.	$?X?Y?b \text{ event}(X, \text{lahir}) \wedge \text{Actor}(X, Y) \wedge \text{Location}(X, b) \wedge \text{place}(b, \text{Jepara}) \wedge \text{object}(Y, \text{Kartini})$
Jepara berada di Jawa Tengah.	$?X?Y?b \text{ event}(X, \text{berada}) \wedge \text{Actor}(X, Y) \wedge \text{Location}(X, b) \wedge \text{place}(b, \text{Jawa}+\text{Tengah}) \wedge \text{object}(Y, \text{Jepara})$

3.2.3. Named Entity and Relation Extraction

The idea is to employ named entity recognizer (NER) and relation extraction (RE). The objects in the semantic representation are the named entity resulted by the NER and the relation between objects are the relation resulted by RE. The relation itself can be extracted from the syntactical parse tree or be trained in a machine learning approach which identifies the relation between named entities. The main problem here is the unavailability of Indonesian relation extraction.

3.3. Selected Method to Generate the Semantic Representation

Based on the appropriateness and component availability, we chose to employ the semantic analysis as the approach. The process in the available semantic analysis includes Indonesian POS Tagger[5], Indonesian PCFG Parser[6], and Indonesian FOL Semantic Analyzer [4].

As an example for further process, the sentence “Kartini lahir di Jepara” will be transformed into FOL of $?X?Y?Z \text{ event}(X, \text{lahir}) \wedge \text{actor}(X, Y) \wedge \text{location}(X, Z) \wedge \text{place}(Z, \text{Jepara}) \wedge \text{object}(Y, \text{Kartini})$ which can be illustrated as shown in Fig. 7.

The text representation resulted by the FOL semantic analysis should be processed further in order to gain the needed representation for the mind map visualization. The illustration of needed representation is shown in Fig. 8.

The process to transform the FOL into Object and Object Relation is by changing the variable into its related value. As for the relation between sentences, the semantic network for one sentence is then joined with the semantic network for other sentences.

Unfortunately, the available Indonesian NLP Tools still give incorrect result due to several things:

1. The error resulted by the Indonesian POS Tagger which then is repaired by adding the lexicon and corpus.
2. The error resulted by the Indonesian PCFG Parser which mainly not working on complex sentences. To fix the error, the required process is to enhance the Indonesian corpus which should be analyzed further.
3. The error resulted by the Indonesian FOL Semantic Analyzer which should be fixed by adding its semantic rule

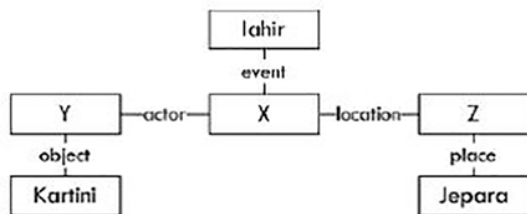


Fig. 7. Illustration of FOL Result for “Kartini lahir di Jepara”



Fig. 8. Illustration of Required Representation for “Kartini lahir di Jepara”

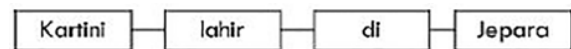


Fig. 9. Example of “Kartini lahir di Jepara” Processed by the Word Order Based Process

To handle the incomplete process caused by the incorrect result of Indonesian NLP Tools, here we added a process to employ the word order. For example, if the sentence “Kartini lahir di Jepara” is failed to be processed by the Indonesian NLP then it will be processed based on its word order and gives result such as shown in Fig. 9.

3.4. Mind Map Visualization

Mind map visualization should follow rules such as drawing main idea in the figure center with branches related with the center. This condition is suitable with radial drawing method which the root as the drawing center can be the main idea while the branches are the entities related with the main idea.

This method still should be modified because the mind map structure resulted from the text is not always a tree. The example of graph of entities resulted from the previous Indonesian sentences is shown in Fig. 10.

The radial drawing method is a variation of layering drawing method [7]. By this, the hierarchical approach which applies layering system can be modified into radial. The illustration is shown in Fig. 11. By this, the entity in the center or the first layer is the main idea of the mind map. Here the main idea is the entity with most relations. The relation of the main idea will placed on the second layer while other entities related with the main idea will be placed in the third layer and so on. After all entities have its layer position, the layer position will be modified into center-distance.

3.5. Interaction in the Mind Map Editor

The mind map representation is usually resulted manually from human mind and the mind map resulted from the automatic generation might give unwanted results, therefore the mind map generator should be completed by its editor. For the mind map editing, we analyzed that there are several things should be handled in the editor:

1. Position of Entity
2. Content of Entity and Relation
3. Structure including the addition or deletion of entity or relation.

4. Color
5. Image addition
6. Size
7. Curvature of Relation Line

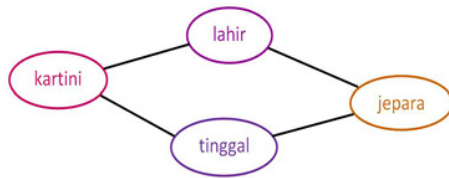


Fig. 10. Graph of Entities of “Kartini lahir di Jepara” and “Kartini tinggal di Jepara”

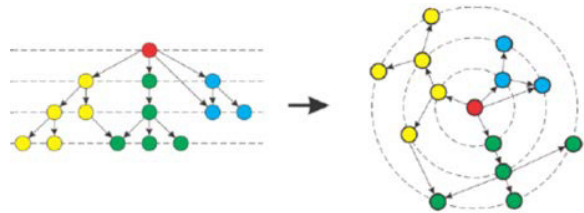


Fig. 11. Radial Tree Drawing as Variation of Layering Tree Drawing

4. Experiment

4.1. Experimental Aim and Data

There are two aims of the experiments: (1) to test the FOL-semantic network transformation; (2) to evaluate the generated mind mapping figure. For the first experiment, we used two types of text: the original text taken from primary school book and the modified text (non-complex sentences). Both texts are then inputted into the mind map generator and the resulted semantic networks are evaluated. There are 10 texts taken from social book 5th grade primary school. For the second experiment, there are 5 respondents were surveyed related with two types of drawing result: the original automatic mind map generator and the modified mind map figure.

4.2. Experimental Result

4.2.1. Experiment to Evaluate the FOL-Semantic Network Transformation

The original text consists of 34 sentences, while the modified text consists of 59 sentences. Here, the complex sentences are modified into simple sentences. Even though the sentences are modified into simple sentences, still not all texts can be processed since the limited rules and training data available in the Indonesian NLP Tools. For the original text, there are only 16 sentences can be processed from 34 sentences which gives accuracy of around 47%. For the modified text, there are 46 sentences can be processed from 59 sentences which gives accuracy of 77%. Mainly the error is caused by the syntactical parser, while the error caused by the transformation is only 1 sentence from both texts.

4.2.2. Experiment to Evaluate Mind Map Drawing Result

Here, we asked 5 respondents to evaluate the legibility of the resulted mind map figure. There are two types of figure: the original automatic one, resulted by the system and the modified one.

As the result, there are 48% of respondents said that the original drawing is readable and easy to understand. As for the modified one, there are 96% respondents said the drawing is readable and easy to understand.

The existence of unimportant words in the drawing makes the drawing difficult to understand. Other reasons for the difficulty are related with the color and the main idea focus.

5. Conclusion

Our mind map generator consists of three components: semantic net generator, mind map visualization and interaction handler. In the semantic network generator, based on the availability of Indonesian NLP tools, we chose to use the available FOL semantic analyzer and added a transformation module to change it into a representation of semantic network. As for the mind map visualization, we used the radial drawing approach. For the root, the node with most relations is chosen as the drawing center. All relations and other objects are connected with the drawing center. The structure is then mapped into radial structure. For the interaction, we defined several things should be handled in the mind map editor. In the experiments, we evaluated the semantic network generator and the mind map visualization. In the semantic network generator, the accuracy achieved was 77% with 1 incorrect sentence (among 59 sentences) resulted by the semantic network transformation. In the mind map visualization, using the original automatic drawing result, there were 48% respondents agree that the drawing result is readable and can be understood. But using the modified drawing result, the result was increased into 96% respondents. The main reason of the illegibility is that the existence of unimportant node in the drawing result.

References

- [1] Buzan, Tony, *Buku Pintar Mind Map*, Jakarta, PT. Gramedia Pustaka Utama; 2005.
- [2] Abdeen, M., El-Sahan, R., Ismaeil, A., El-Harouny, S., Shalaby, M., Yagoub, M. C. E. Direct Automatic Generation of Mind Maps from Text with M2Gen. In *Proceeding of IEEE Toronto International Conference Science and Technology for Humanity*, 2009; p. 95-99, Toronto, Canada.
- [3] Brucks, C., Schommer, C. Assembling Actor-based Mind-Maps from Text Streams. Master Thesis, University of Luxembourg, Department of Computer Science and Communication; 2008.
- [4] Ferdian, F., Purwarianti, A. Implementation of Semantic Analyzer in Indonesian Text-Understanding Evaluation System. In *Proceedings of IEEE International Conference on Computational Intelligence and Cybernetics*, Bali; 2012.
- [5] Wicaksono, A. F., Purwarianti, A. HMM based Part of Speech Tagger for Bahasa Indonesia. In *Fourth International MALINDO Workshop*, Jakarta; 2010.
- [6] Afif, I. Studi Perbandingan Kinerja Algoritma CYK dan Algoritma Earley pada Pengurai Kalimat Menggunakan Probablistic Context Free Grammar Bahasa Indonesia Sederhana. Final Project of Undergraduate, Bandung Institute of Technology; 2011.
- [7] Batista, G., Eader, P., Tamassia, R., Tollis, I. *Graph Drawing: Algorithm for the Visualization of Graphs*, Prentice Hall; 1999.