Jurnal Linguistik Komputasional
Indonesian Association for Computational Linguistics
JLK
Vol 2, No 1
Maret 2019

# Employing Dependency Tree in Machine Learning Based Indonesian Factoid Question Answering System

Irfan Afif [#1], Ayu Purwarianti [#2]

[#] *School of Electrical and Informatics Engineering, Bandung Institute of Technology Bandung, Indonesia*
[1]irfan.apip@gmail.com
[2]ayu@stei.itb.ac.id

***Abstrak*— We proposed the usage of dependency tree information to increase the accuracy of Indonesian factoid question answering. We employed MSTParser and Universal Dependency corpus to build the Indonesian dependency parser. The dependency tree information as the result of the Indonesian dependency parse is used in the answer finder component of Indonesian factoid question answering system. Here, we used dependency tree information in two ways: 1) as one of the features in machine learning based answer finder (classifying each term in the retrieved passage as part of a correct answer or not); 2) as an additional heuristic rule after conducting the machine learning technique. For the machine learning technique, we combined word based calculation, phrase based calculation and similarity dependency relation based calculation as the complete features. Using 203 data, we were able to enhance the accuracy for the Indonesian factoid QA system compared to related work by only using the phrase information. The best accuracy was 84.34% for the correct answer classification and the best MRR was 0.954.**

***Keywords*—question answering, factoid, machine learning, dependency tree**

## I. INTRODUCTION

Question Answering (QA) system is part of Natural Language Processing that takes question in natural language and generate an answer from available corpus automatically. In general, QA system consists of three components, which are Question Analyzer, Passage Retriever and Answer Finder [1]. Question Analyzer analyzes input question and extract information needed to find the answer. Passage Retriever extracts relevant documents and paragraphs. Answer Finder extracts the answers from documents or paragraph given by Passage Retriever. The techniques used in QA can be classified into heuristics based and machine learning based. In both techniques, one can employ various text information, ranging from lexical information to semantic information contained in the input question, candidate answer paragraph or sentence.

Indonesian is a low resource language. High accuracy natural language tool for Indonesian haven't available yet, especially for syntactic [3][4] and semantic parser [5]. QA system for Indonesian language has been built using the available tools such as Purwarianti [6] using machine learning approaches with lexical features, Larasati & Manurung [7] using heuristic approaches with semantic representation, and Zulen & Purwarianti [2] employed phrase based features to enhance the lexical features in machine learning approach. Even though there is a research on Indonesian QA using a semantic analyzer, but there is no result reported on the experiments and the available Indonesian semantic analyzer still has a low accuracy [5]. Meanwhile, the research on Indonesian dependency parser [10] has been developed with higher accuracy than the constituent parser [17] using existing dependency parser such as MaltParser [8] and MSTParser [9]. Even though there is a research on using a dependency tree on the QA system for English speech corpora [11], but the accuracy was still low. The previous Indonesian QA system with machine learning technique was able to increase the accuracy using phrase information [2] compared to lexical based one [16]. One of the weaknesses in this previous Indonesian QA system is occurrences of ambiguity answer. Here we proposed the usage of dependency tree information to handle such problem.

## II. INDONESIAN DEPENDENCY PARSER

Dependency parser aims to yield dependency syntax tree to a given input sentence. Different with a constituent syntax tree, the dependency tree doesn't contain phrase information. The example of a dependency tree is shown in the figure below [Fig. 1]. Here, words are connected one another as a head and dependent relation. One word can only have one head and can have many dependents.

There are two methods in dependency parser: grammar-driven method and data-driven method. Indonesian dependency grammar is not yet available. Kamayani and

Purwarianti [4] developed Indonesian dependency grammar based on Stanford Dependency Label. The research result is limited to parsing simple sentences only.

The other method is a data-driven one. Rahman [10] compared the dependency parsing algorithms already available in MaltParser and MSTParser, while Green [15] conducted ensemble learning for several dependency parsing algorithms. The experimental result of Rahman [10] can be seen in Table 1. Since our focus here is in the QA system, to minimize the execution time of the system, we decided to use a single dependency parsing algorithm to yield the dependency tree information. Here, we chose to use Chu-Liu-Edmonds algorithm [13][14] as our dependency parser, which achieved the highest accuracy in previous experiments (Rahman [10]).

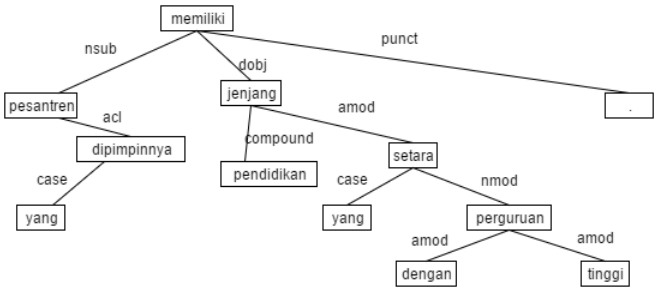


Fig. 1. Dependency tree example

TABLE I
EXPERIMENTAL RESULT OF DEPENDENCY PARSER (RAHMAN[10])

| **Dependency Parser** | **Algorithm** | **Accuracy** |
|---|---|---|
| MaltParser | Nivre-eager | 83,5% |
| | Nivre-standard | 82,9% |
| | Covington projective | 82,4% |
| | Covington non-projective | 82,6% |
| | Stack projective | 83,3% |
| | Stack eager | 83,7% |
| | Stack lazy | 83,9% |
| | Planar | 84,1% |
| | 2-Planar | **84,7%** |
| MSTParser | Eisner | 85,8% |
| | Chu-Liu-Edmonds | **86,1%** |

In the experiment, Rahman built their own Indonesian dependency corpus with 2,098 sentences. Unfortunately, there is no label information on the edge between term nodes. In our analysis, the label information on edge is important in verifying whether a term is an answer or not. To address this problem, we used Universal Dependency which has label dependency on it. Indonesian Universal Dependency corpus contains general POSTag (such as Noun and Verb), yet it does not have specific POSTag (such as verb transitive: VBT and verb intransitive: VBI). Specific POSTag can be added using Indonesia POSTagger [12]. In this experiment we used Chu-Liu-Edmonds algorithm in MSTParser with Universal Dependency corpus added with specific POSTag.

## III. PHRASE BASED QUESTION ANSWERING SYSTEM

In previous experiments, Zulen & Purwarianti[2] developed Question Answering System using machine learning approach based on keyword and phrase structure. Question Analyzer module classified EAT (Expected Answer Type) using rules based on interrogative words and clue words. Lexical and phrase information extracted using POSTagger, stemmer, and phrase chunker. Passage Retriever module build using Lucene [18]. Answer Finder module selects answer candidates using NE Tagger. Lexical and phrase information is used as features in machine learning. Feature extraction is enhanced using reference resolution technique. Three machine learning algorithms are used to rank answer candidates, namely Support Vector Machine, Maximum Entropy and J48. The experiments used 203 sentences consist of five EATs which are person, organization, location, datetime and quantity. The experimental result can be seen in Table 2 and Table 3. Table 2 accuracy calculated based on precision of class answer-Yes and class answer No in machine learning prediction. Precision calculated using equation (1). MRR in Table 3 calculated using equation (2) where Q is total question used and rank is the ranking of the answer of the question.

$$Accuracy = \frac{True\ positive\ of\ class\ answer-Yes}{Totan\ of\ class\ answer-Yes} \text{x } 100\% \quad (1)$$

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{Q} \frac{1}{rank_i} \quad (2)$$

TABLE III
ANSWER CLASSIFICATION ACCURACY [2]

| EAT | SVM | | J48 | | MaxEnt | |
|---|---|---|---|---|---|---|
| | Yes | No | Yes | No | Yes | No |
| Person | 60.2 | 82.3 | 42.4 | 85.1 | 63.8 | 83.5 |
| Organization | 62 | 89.0 | 42.9 | 90.8 | 62.7 | 86.3 |
| Location | 74.4 | 61.8 | 50.6 | 80.6 | 75.6 | 64.8 |
| Datetime | 72.8 | 70.8 | 46.5 | 86.3 | 81.2 | 74.0 |
| Quantity | 78.1 | 67.1 | 66.8 | 78.3 | 76.8 | 70.0 |
| All | 66.8 | 75.8 | 47.2 | 84.7 | 69.1 | 76.9 |

TABLE IIIII
MRR VALUE FOR ANSWER FINDER [2]

| EAT | Baseline | SVM | J48 | MaxEnt |
|---|---|---|---|---|
| Person | 0.586 | 0.846 | 0.635 | 0.854 |
| Organization | 0.801 | 0.786 | 0.663 | 0.797 |
| Location | 0.696 | 0.813 | 0.614 | 0.716 |
| Datetime | 0.849 | 0.757 | 0.573 | 0.799 |
| Quantity | 0.680 | 0.811 | 0.601 | 0.755 |
| All | 0.723 | 0.802 | 0.617 | 0.784 |

Further analysis showed that the answer ambiguity problem was occurring in Zulen & Purwarianti experiments [2]. Answer ambiguity problem occurs when a wrong answer is ranked the same with highest ranked correct answer. Example of this case can be seen in Fig. 2. Answer ambiguity problem cannot be measured with MRR. However, this problem should be considered when measuring QA system accuracy. We cannot state QA system able to identify right answer as ranked first when there is wrong answer included in the same rank.

=========ANALISIS PERTANYAAN=========
Pertanyaan : Pada tanggal berapa Hari Kesaktian Pancasila diperingati ? (When the Day of Kesaktian Pancasila celebrated ?)
EAT: DATETIME
Kata Kunci: Hari, Kesaktian, Pancasila, diperingati

Pohon Dependensi:
Pada : 2 tanggal - case
tanggal : 7 diperingati - nmod
berapa : 4 Hari - nummod
Hari : 7 diperingati - nsubjpass
Kesaktian : 4 Hari - compound
Pancasila : 5 Kesaktian - name
diperingati : 0 - root
? : 7 diperingati - punct

=========HASIL JAWABAN=========
1. **30 September**
Kalimat : Maka **30 September** 1965 diperingati sebagai Hari Peringatan Gerakan 30 September G30S-PKI dan tanggal 1 Oktober ditetapkan sebagai Hari Kesaktian Pancasila, memperingati bahwa dasar Indonesia, Pancasila, adalah sakti, tak tergantikan .
(30 September 1965 is celebrated as the Day of G30S-PKI and 1 October is celebrated as the Day of Kesaktian Pancasila, )
1. **1 Oktober**
Kalimat: Maka 30 September 1965 diperingati sebagai Hari Peringatan Gerakan 30 September G30S-PKI dan tanggal **1 Oktober** ditetapkan sebagai Hari Kesaktian Pancasila, memperingati bahwa dasar Indonesia, Pancasila, adalah sakti, tak tergantikan .
(30 September 1965 is celebrated as the Day of G30S-PKI and 1 October is celebrated as the Day of Kesaktian Pancasila, )

Fig. 2. Answer Ambiguity Problem Example

This problem occurred in previous research [2] when two or more answer candidates have same the value in all machine learning features used in the experiment. It often occurred between two candidate answers in the same sentence. It shows that word based and phrase based features is not enough to differentiate answers.

## IV. DEPENDENCY TREE BASED ANSWER FINDING METHOD

Two dependency based answer finding methods are developed in this research. The first method employs dependency tree based features in machine learning. The second method uses distance in dependency tree with heuristic approach.

TABLE IVV
DEPENDENCY TREE BASED FEATURES

| No | Feature | Definition |
|---|---|---|
| 1 | countRootVerb | Occurrences of verb which has position as root or dependent to root in question dependency tree.<br>Example in Fig 2: celebrated (diperingati) as RootVerb |
| 2 | countDependencyParent Sentence | Ratio of single keyword which has the same parent in question with all keywords occurrences in sentence.<br>Example in Fig 2: dependency relation between Kesaktian and Pancasila |
| 3 | countDependencyParent Paragraph | Ratio of single keyword which has the same parent in question with all keywords occurrences in paragraph.<br>Example in Fig 2: dependency relation between Kesaktian and Pancasila |
| 4 | countDependencyLabel ParentSentence | Ratio of single keyword which has the same parent and label in question with all keywords occurrences in sentence.<br>Example in Fig 2: dependency relation with label name between Kesaktian and Pancasila |
| 5 | countDependencyLabel ParentParagraph | Ratio of single keyword which has the same parent and label in question with all keywords occurrences in paragraph.<br>Example in Fig 2: dependency relation with label name between Kesaktian and Pancasila |
| 6 | countDistanceMainVerb | Ratio between 1 and average of answer distance with verb keywords in dependency tree.<br>Example in Fig 2: Average of distance in dependency tree between words in answer with word ‘celebrated (diperingati)’. |
| 7 | counDistanceQuery | Ratio between 1 and average of answer distance with noun keywords in dependency tree. |
| 8 | counDistanceQueryMin | Ratio between 1 and average of answer distance with verb keywords in dependency tree. If there are occurrences of two or more identical keyword, then used the nearest keyword. |

Dependency based feature used in this experiment can be categorized into three groups. We used dependency relation similarity between keyword in question and in sentence answer with and without label as group one. Distance calculation between keyword and candidate answer as group two. Keyword position in dependency tree as group three. Table 4 contains dependency tree based feature in this experiment. Table 5 contains grouping of previous experiment features and dependency tree features proposed in this experiment.

Heuristic method using distance in dependency tree is used to differentiate between answer in the same sentence. This method is used together with machine learning methods. Answers are ranked using machine learning. If an answer ambiguity problem occurred, then we ranked answers by distance between answer and keyword in sentence dependency trees.

TABLE V
GROUP FEATURES USED IN EXPERIMENT

| No | Group | Feature |
|---|---|---|
| **1** | Keyword | keywordSentence |
| | | keywordParagraph |
| **2** | KeywordPhrase | phraseKeywordSentence |
| | | phraseKeywordParagraph |
| | | phraseSimilaritySentence |
| | | phraseSimilarityParagraph |
| **3** | PhraseDistance | phraseDistanceAll |
| | | phraseDistanceVerb |
| | | phraseDistanceNoun |
| **4** | PhraseType | countVerbPharse |
| | | countNounPharse |
| **5** | DependencyLocation | countRootVerb |
| **6** | DependencySimilarity | countDependencyParent Sentence |
| | | countDependencyParent Paragraph |
| | | countDependencyLabel ParentSentence |
| | | countDependencyLabel ParentParagraph |
| **7** | DependencyDistance | counttDistanceMainVerb |
| | | countDistanceQuery |
| | | countDistanceQueryMin |

We analyze combined feature based on analysis of subset and scope of the features. There are cases where keyword does not appear in the sentence which can be seen in Fig. 3. Feature keywordSentence is eliminated based on that case. Feature countRootVerb is subset of feature keywordSentence, hence feature countRootVerb is also eliminated. Phrase based feature used in sentece only, hence feature phraseKeywordParagraph and feature phraseSimilarity Paragraph is eliminated. Distance based featured is eliminated if the heuristic method based on distance in dependency tree is used. The eliminated feature can be seen in Table 6.

| **Question : Siapa orang Afrika-Amerika pertama yang terpilih sebagai presiden Amerika Serikat ?**<br><br>**Keywords: orang, Afrika-Amerika, pertama, terpilih, presiden, Amerika, Serikat**<br>**Count Keyword : 7** | |
|---|---|
| **Right Answer: Barack Obama**<br>**Sentence: Barack Obama akan menjadi presiden separo Afrika-Amerika pertama .**<br><br>**Keywords: presiden, afrika-amerika, pertama**<br>**Count Keyword: 3** | Wrong Answer: Martin Van Buren<br>Sentence: Pada 4 Maret 1837 Martin Van Buren terpilih menjadi Presiden Amerika Serikat yang ke-8 .<br><br>Keywords: terpilih, presiden, Amerika, Serikat<br>Count Keyword: 4 |

Fig. 3. Wrong answer has more keywords than right answer case example

TABLE VI
THE RESULT OF ELIMINATED FEATURE IN COMBINED FEATURE FOCUS ON SENTENCE

| No | Feature | Reason |
|---|---|---|
| **1** | keywordSentence | Contradicting case. Word calculation not used in sentence |
| **2** | phraseKeywordParagraph | Phrase calculation not used in paragraph |
| **3** | phraseSimilarityParagraph | phrase calculation not used in paragraph |
| **4** | phraseDistanceAll | If heuristic method used |
| **5** | phraseDistanceVerb | If heuristic method used |
| **6** | phraseDistanceNoun | If heuristic method used |
| **7** | countVerbPharse | Subset of feature keywordSentence |
| **8** | countNounPharse | Subset of feature phraseKeywordSentence |
| **9** | countRootVerb | Subset of feature phraseKeywordSentence |
| **10** | counttDistanceMainVerb | If heuristic method used |
| **11** | countDistanceQuery | If heuristic method used |
| **12** | countDistanceQueryMin | If heuristic method used |

## V. EXPERIMENT

### *A. Indonesian Dependency Parser Accuracy*

Indonesian Dependency parser built using Chu-Liu-Edmonds algorithm in MSTParser with Universal Dependency training data corpus and tested using Universal Dependency testing data corpus. Testing consists of two scenarios. Scenario one using Universal Dependency as it is and scenario two using Universal Dependency added with component POSTAG, Universal Dependency added with component POSTAG able to achieve higher accuracy. Experiment result can be seen in Table 7.

TABLE VII
INDONESIAN DEPENDENCY PARSER EXPERIMENT RESULT

| No | Measurement | Universal Dependency (Scenario 1) | Universal Dependency with component POSTAG (Scenario 2) |
|---|---|---|---|
| **1** | Dependency accuracy | 79,31% | 81,29% |
| **2** | Dependency accuracy with label | 60,68% | 74,32% |

### *B. QA System Accuracy*

We use experimental data from previous experiment [2] using 203 questions categorized into five different EAT (Expected Answer Type). This data is split with ratio 75 : 25 into training data and testing data. Machine learning used in the experiment also based on previous experiment [2], which use Support Vector Machine (SVM), Maximum Entropy and J48 algorithm. Testing scenario is arranged based on different combinations of machine learning features and dependency tree heuristic approach. Features used in the scenario is shown in Table 8. Scenario experiment is described in Table 9.

TABLE VIII
FEATURE USED IN EXPERIMENT

| No | Name | Using Feature |
|---|---|---|
| 1 | Feature1 | Keyword + Phrase (baseline) |
| 2 | Feature2 | Dependency Tree |
| 3 | Feature3 | Keyword + Phrase + Dependency Tree |
| 4 | Feature4 | Combined feature focus on sentence |

TABLE IX
SCENARIO EXPERIMENT

| No | Name | Feature | Heuristic | Algorithm |
|---|---|---|---|---|
| 1 | Scenario1 | Feature1 | No | MaxEnt |
| 2 | Scenario2 | Feature1 | Yes | MaxEnt |
| 3 | Scenario3 | Feature2 | No | MaxEnt |
| 4 | Scenario4 | Feature2 | Yes | MaxEnt |
| 5 | Scenario5 | Feature3 | No | SVM |
| 6 | Scenario6 | Feature3 | Yes | SVM |
| 7 | Scenario7 | Feature4 | No | SVM |
| 8 | Scenario8 | Feature4 | Yes | SVM |

The experiment is performed to measure machine learning classifying accuracy, using feature from previous research, features based on dependency tree and heuristic method proposed in this experiment. Result of the experiment can be seen in Table 10 and Table 11. Combined features based on analysis of redundancy and scope with a heuristic method based on dependency tree able to achieve the highest score. The highest scored achieved in this experiment is better than the highest score in previous method. This indicates that feature combination using word based, phrase based and dependency based calculation and heuristic method calculation able to answer factoid question better and address problems in previous research.

TABLE X
ANSWER CLASSIFICATION ACCURACY

| Feature | SVM | | J48 | | MaxEnt | |
|---|---|---|---|---|---|---|
| | Yes (%) | No (%) | Yes (%) | No (%) | Yes (%) | No (%) |
| Feature 1 | 75.10 | 77.53 | 47.39 | 78.95 | 72.29 | 76.85 |
| Feature 2 | 58.23 | 67.91 | 28.51 | 81.38 | 59.04 | 70.68 |
| Feature 3 | 70.68 | 79.63 | 31.73 | 86.42 | 70.68 | 79.57 |
| Feature 4 | 79.92 | 75.16 | 73.49 | 77.36 | 65.86 | 75.38 |

TABLE XI
MRR VALUE AND AMBIGUITY ANSWER OCCURRENCES FOR EACH SCENARIO

| No | Scenario | MRR | Ambiguity Answer |
|---|---|---|---|
| 1 | Scenario1 | 0.9114 | 19 |
| 2 | Scenario2 | 0.8682 | 0 |
| 3 | Scenario3 | 0.5862 | 0 |
| 4 | Scenario4 | 0.5862 | 0 |
| 5 | Scenario5 | 0.8872 | 0 |
| 6 | Scenario6 | 0.8872 | 0 |
| 7 | Scenario7 | 0.9774 | 38 |
| 8 | Scenario8 | 0.9536 | 0 |

From the test result, there are cases where QA system failed to score right answer as the number one answer. The cause of this problem is:

- NE Tagger misses tagging location answer.
- Similarity between question and wrong answer candidate is higher than similarity between question and right answer candidate.

## VI. CONCLUSION

In this research, a statistical based QA system using dependency tree information for Indonesian has been made. Dependency tree is used to increase accuracy and solved problem in phrase-based method used in previous research. Indonesian dependency parser is built using

Chu-Liu-Edmonds algorithm in MSTParser with Universal Dependency corpus. Dependency tree used in answer finding method using machine learning approach and heuristic approach.

Machine learning approach consists of word based feature, phrase based feature, and dependency based feature. Heuristic approach based on distance in dependency tree. By combining machine learning approach and heuristic approach, QA system builds in experiment able to perform better than previous research method and able to eliminate ambiguity answer problem. There still questions that cannot be answered by this method. Another method can be applied to solve this problem such as using semantic approach.

## REFERENCES


[1] Harabagiu, S. M., M. A. Pasca and S. J. Maiorano. "Experiments with Open-Domain Textual Question Answering," Proceedings of the 18th International Conference on Computational Linguistics (COLING), Saarbruken, Germany, 2000.

[2] A. A. Zulen and A.Purwarianti, "Using phrase-based approach in machine learning based factoid Indonesian question answering," Proceedings of CISAK, 2013.

[3] I. Afif, "Study And Performance Comparison Of Cyk Algorithm And Earley Algorithm In Parser Using Simple Indonesian Probabilistic Context Free Grammar," Bachelor's Program Thesis, Institut Teknologi Bandung, 2011.

[4] M. Kamayani and A. Pirwarianti, "Dependency Parsing for Indonesian with GULP," in Proceeding of The 3rd ICEEI, 2011.

[5] F. Ferdian and A. Purwarianti, "Implementation of Semantic Analyzer in Indonesian Text-Understantding Evealuation System," in Proceeding of CYBERNETICSCOM, 2012.

[6] A. Purwarianti, M. Tsuchiya and S. Nakagawa. "A Machine Learning Approach for Indonesian Question Answering System," Proceedings of the 25th IASTED International Multi-Conference: Artificial Intelligence and Applications, Innsbruck, Austria, 2007.

[7] S.D. Larasati, dan R. Manurung, "Towards a Semantic Analysis of Bahasa Indonesia for Question Answering," Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics, Melbourne, Australia, 2007.

[8] Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., . . . Marsi, E., "MaltParser: A Language-Independent System for Data-Driven Dependency Parsing," Natural Language Engineering. 13, hal. 95-135. Cambridge Univ Press, 2007.

[9] McDonald, R., Pereira, F., Ribarov, K., dan Hajič, J., "Non-Projective Dependency Parsing Using Spanning Tree Algorithms," Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (hal. 523-530). Association for Computational Linguistics, 2005.

[10] A. Rahman, "Pemanfaatan Teknik Ensemble Untuk Pengurai Dependensi Bahasa Indonesia," Bachelor's Program Thesis, Institut Teknologi Bandung, 2011.

[11] Comas, P. R., Turmo, J., Marquez, L., "Using dependency parsing and machine learning for factoid question answering on spoken documents," *11th Annual Conference of the International Speech Communication Association*, Makuhari, Chiba, Japan, September 26-30, 2010.

[12] A.F., Wicaksono and A., Purwarianti, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia," *Proceedings of the 4th-International MALINDO Workshop (MALINDO2010)*, Jakarta, Indonesia, 2010.

[13] Chu, Y.-J., and Liu, T.-H. "On Shortest Arborescence of a Directed Graph", *Scientia Sinica*, 1396, 1965.

[14] Edmonds, J., "Optimum Branchings". *Journal of Research of the National Bureau of Standards B*, 233-240. 1967.

[15] Green N, Larasati S and Žabokrtský Z, "Indonesian dependency treebank: Annotation and parsing," *In: Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation*, Faculty of Computer Science, Universitas Indonesia, Bali,Indonesia, pp 137–145, 2012.

[16] A. A. Zulen and A. Purwarianti, "Study and Implementation of Monolingual Approach on Indonesia Question Answering for Factoid and Non-Factoid Question," *in Proceedung of The 25th PACLIC*, 2011.

[17] A Purwarianti, A Saelan, I Afif, F Ferdian, A. F. Wicaksono, "Natural Language Understanding Tools with Low Language Resource in Building Automatic Indonesian Mind Map Generator," *International Journal on Electrical Engineering and Informatics* 5 (3), 256, 2013.

[18] M. McCandles, E. Hatcher and O. Gospodnetic, "*Lucene in Action*," 2nd ed,. USA: Manning Publication, 2010.