Journal of Engineering Science and Technology
Special Issue on ICIT2022 Vol. 19, No. 1 (2024) 1 - 11

TEXT-BASED TAXONOMY LEARNING FOR THE INDONESIAN LANGUAGE USING MACHINE LEARNING

MISINEM[1], ERMATITA[2], DIAN PALUPI RINI[2], TRI BASUKI KURNIAWAN[3], DESHINTA ARROVA DEWI[4,*]

[1]Faculty of Vocational, University of Bina Darma, Palembang, Indonesia
[2]Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia
[3]Faculty of Computer Science, University of Bina Darma, Palembang, Indonesia
[4]Faculty of Data Science and Information Technology, INTI International University, Malaysia
*Corresponding Author: misinem@binadarma.ac.id

Abstract

According to taxonomy, animals are classified as mammals, reptiles, or crocodiles. This biological taxonomy defines the similarities, differences, and relationships between animals. Internet scientists and engineers 'borrow' the concept and function of biological taxonomy when developing taxonomies for the Internet. Developing taxonomies for the Internet by hand, like biological taxonomy, is complex and costly because the task takes time and requires field ingenuity. Thus, computer scientists have used artificial intelligence (AI) approaches to generate taxonomies from the text automatically. Machine learning algorithms enable the machine to 'read' the text and then 'learn' to construct a taxonomy based on its context. The primary goal of this research is to create a more effective taxonomic learning algorithm from Indonesian language text than the existing hybridization algorithms. This study investigates the effectiveness of hybrid algorithms between the Bisection K-Means Algorithm (BKMA) and the hybrid algorithm called the Bisection-Firefly Algorithm (BFA). Data from experiments on three Indonesian language texts from the Biochemistry, Information Technology, and Islamic Jurisdiction are gathered in this empirical study. When dealing with data sparseness problems, a comparison of accuracy using the F-measure reveals that the BFA constructs more accurate taxonomies, giving more effective and robust results than another algorithm. However, this exploratory study needs to be expanded with a larger Indonesian language corpus to test the algorithm's robustness and resilience when dealing with a more general corpus rather than its technical and specific corpus of texts. The syntactic dependency-based extraction technique must be improved because it has resulted in severe data sparsity. As a result, it creates a new challenge for researchers studying taxonomic learning from Indonesian language texts.

Keywords: Education, Features, Firefly algorithm, Indonesian language, Learning opportunity. Machine learning, Taxonomy learning.

1. Introduction

Taxonomy is a scientific proposal for classification, such as the classification of organisms. The scientist explains the taxonomic relationship between one organism and another. Animals, for example, can be divided into several groups, such as mammals, reptiles, and crocodiles. This biological taxonomy creates a classification scheme that allows scientists to define similarities, differences, and relationships between animals. Cats and elephants are examples of taxonomic relationships.

Internet scientists and engineers use biological taxonomy concepts and functions when developing taxonomies for the Internet. A taxonomy for the Internet is a collection of words and word specifications, such as the type of word relationship with other words, that are organized in a hierarchy and can be used to map keywords to relevant resources based on meaning or semantics [1]**.** Manually developing a taxonomy, like biological taxonomy, is neither simple nor inexpensive. Creating a practical taxonomy this task takes time and requires domain experts.

Thus, computer scientists have used artificial intelligence (AI) approaches to create taxonomies automatically from the text. Taxonomy learning is the use of machine learning approaches to develop taxonomies automatically. Taxonomy learning algorithms are intended to enable machines to 'read' text and then 'learn' to construct a taxonomy based on the context derived from the text. The context in this study refers to learning the characteristics of 'verbs,' which are frequently used in conjunction with 'nouns.' A syntactic dependency-based feature acquisition method is used to acquire these features. However, this method may result in data sparsity issues [2]. The problem of sparse data means that the characteristics (verbs) for a noun that are the basis of the context so that the noun's position in a 'concept hierarchy' can be determined are insufficient.

Recent research has proposed a new technique for automatically developing taxonomies from English texts [2-4]. However, the developed methods and techniques are only applicable to English-language texts. According to Wang et al. [5], taxonomy learning research from texts other than English is worth investigating due to language structure and style differences. Machine learning techniques that have previously been shown to be effective in English texts are frequently tested in research questions. Is the method applicable to texts in other languages, such as Bahasa (Indonesian Language)? Data scarcity is one of the issues that often stymies taxonomy learning. Many factors can contribute to data sparsity, extracting features (words) from the text (corpus).

According to Cimiano [2], the attribute value for an extracted term may be incorrect or suffer from rarity issues due to the following reasons: natural language processing tools fail to label word types accurately. As a result, not all words extracted from syntactic dependencies are correct, and not all syntactic dependencies obtained (even if correct) will help distinguish between an object and different objects. Another problem is the assumption that information completeness will never be met [6].

Research in this field is frequently focused on developing more effective taxonomy learning algorithms from the text than existing ones. To address the issue of noise and data sparsity, Cimiano [2] used algorithms such as Formal Concept Analysis, GAHC, Agglomerative Clustering, and K-Mean Bisection. According to

Wang et al. [5] taxonomy learning methods from the text can be divided into three categories, namely:

- extraction of taxonomic relationships based on Language patterns [7]
- the method is based on the distribution (characteristics)
- taxonomy development from the sentence containing the taxonomy keyword, which is 'is.' For example, 'cats are mammals.

The second category, distribution-based taxonomy learning, is the focus of this research. This method is an example of unsupervised machine learning. This method is based on Harris's [8] distribution hypothesis to create a technique for identifying synonyms, concepts, and taxonomic relationships. According to [8], "words are the same if they share a similar context." The concept of 'situational context' introduced the nature of context-dependence [9]. Firth's statement that "you will know a word by the other words associated with it" has become essential in text mining research, information access, and ontology learning [9]. As a result, the following theory underpins this study:

- The Contextual (distribution) meaning hypothesis [8, 9] states that their use in the text determines the meaning of words.
- According to the semantic similarity context hypothesis by Hadj Taieb et al. [10], words with context similarity also have a semantic similarity.

Data collected by Ellis [11] support the claim that humans summarize the context representation of a word based on their experience with multiple linguistic contexts. These findings lend credence to the meaning context hypothesis. Several studies have been conducted to test the hypothesis's validity, including those of Casillas et al. [12] and Ellis [11]. Their empirical investigation has confirmed the above hypothesis's fact.

Most studies using this method are based on distribution similarity measures such as cosine, Jaccard, Jensen-Shannon divergence, or the LIN. This approach measures the relationship between words when forming a concept hierarchy Wang and Dong [13] used a non-Euclidean similarity measure. In contrast, Berner et al. [14] used an artificial neural network based on the distribution hypothesis, and Bi et al. [15] used a graph-based method. Literature review shows that Qiu et al. [16] still used this method and designed a taxonomy learning method for Chinese texts.

Some previous studies, such as [2], focused on solving the issue of noise and data sparseness. Cheng et al.[17] describe some characteristics that an algorithm should possess, one of which is the robustness or durability of the algorithm to produce a quality taxonomy still when 'served' with complex data.

Therefore, there is a need to develop a robust taxonomy learning algorithm that can generate a quality taxonomy even if the data obtained suffers from a severe rarity problem. This paper discusses how the Bisection K-Means Algorithm (BKMA), and Bisection-Firefly Algorithm (BFA) can form a taxonomy from Indonesian text.

This paper begins with an introduction to the problem of this research and previous research. The following section explains the research methodology used in this study. This section also discusses the experimental setup, and the developed technique is described in detail. After the proposed algorithm is concerned,

experimental results will be presented and discussed in the Results and Discussion section. Then, the analysis of the results and the comparison of the results between the two algorithms are displayed. The performance evaluation of the technique is reported as well in this section.

2. Methodology

Three different Indonesian language texts were used to ensure that the results obtained were significant and not by chance to test the robustness of the proposed method. The text and data set used in this study were created in collaboration with domain experts from each domain (i.e., Biochemistry, Information Technology, and Islamic Jurisdiction).

The gold standard for each domain (text) is compared to the taxonomy quality obtained from the data set extracted using the feature extraction method based on the syntactic dependency method [2]. The comparative taxonomies developed by domain experts based on the texts used are detailed in Tables 1 and 2.

Table 1. Gold standard and comparison.

	Biochemistry	Information Technology	Islamic Jurisdiction
Number of concepts	164	478	422
Number of leaves	113	275	393
Average Depth	3.34	2.08	2.26
Maximum depth	7	8	6
No. maximum child	82	96	34
Average child	7.06	4.82	5.46

Table 2. Gold standard and comparison.

	Biochemistry	Information Technology	Islamic Jurisdiction
Total hypernym/hyponym	2	119	52
Sum of contextual features	80	137	162
Maximum number of features for a term	6	61	52
The minimum number of features for a term	1	1	1
Min (Number of features)	1.51	2.08	2.47
% terms with only one (1) feature	76.77%	68.57%	62.93%

2.1. K-Means algorithm

The K-Means algorithm is an iterative algorithm that attempts to partition the dataset into k distinct non-overlapping subgroups (clusters), with each data point belonging to only one group. It tries to keep intra-cluster data points as similar as possible while keeping clusters as different (far) as possible. It assigns data points to clusters so that the sum of the squared distances between the data points and the cluster's centroid (the arithmetic mean of all the data points in that cluster) is as small as possible. The lower the variation within clusters, the more homogeneous (similar) the data points within the cluster [18].

An improved K-means clustering algorithm can be applied to large amounts of data while maintaining acceptable precision rates [18]. The distances from one point to its two nearest centroids and their variations were used in the last two iterations of this method. Points with equal distance thresholds more significant than the equal distance index were excluded from distance calculations and clustered. Even though these points are compared with the research index —cluster radius—again in the algorithm iteration, the excluded points are included in the calculations if their distances from the stabilized cluster centroid are more significant than the cluster radius. It has the potential to improve clustering quality.

2.2. Bisection K-means algorithm

The bisecting K-means clustering technique is a minor modification to the standard K-Means algorithm that fixes the procedure for dividing data into clusters. As with K-means, we begin by initializing k centroids (You can either do this randomly or can have some prior). Following that, we use regular K-means with k=2. This bisection step is repeated until the desired number of clusters is reached. After the first Bisection (when there are two clusters) is finished, several strategies for selecting one of the clusters and re-iterating the entire bisection and assignment process within that cluster. For example, choose the cluster with the highest variance or the spread-out cluster, select the cluster with the most data points, and so on [19].

2.3. Firefly algorithm

According to Lewis and Cratsley [20], fireflies (Coleoptera: Lampyridae) are among the most 'charismatic' insects, mainly because of the captivating way fireflies have inspired scientists. According to Jaikla et al. [21], there are significant differences in the nervous systems of male and female fireflies, implying that these insects have different control functions. Fireflies are a type of insect that belongs to the beetle family. There are over 2000 species of fireflies in the world.

There are several variations of the firefly algorithm in the literature. Fister et al. [22] proposed a classification scheme for categorizing the Firefly Algorithm based on its parameter settings. This Firefly Algorithm parameter setting is critical for good performance and should be carefully chosen [23]. In general, there are two approaches to correctly setting the algorithm parameters. The first method is to tune the parameters before running the algorithm, and the second is to adjust the parameters after running the Firefly Algorithm. After completing an iteration, parameter tuning is performed. Apart from being based on parameter setting methods, the classification used by Fister et al. [22] also considers the components and features of the Firefly Algorithm.

2.4. Bisection-firefly algorithm

The main difference between BFA and the original FA is that the Bisection implements the hierarchical formation of BFA clusters because the original FA was not designed to form conceptual hierarchies or taxonomies. The following Algorithm 1 displays the BFA pseudocode.

Pseudocode of Bisection-Firefly Algorithm

1: Input: data (nouns and features) extracted from the text.

2: Derive: number of iterations, best solution, attractiveness, clustering threshold value (t = 100; s *=* ∅; γ = 1.0; p = 0.05; *PopulationNum* = 50)
3: $P(0)$ = I*nitialFireflyPopulation*();
Generate a population of 50 Fireflies. A Firefly has as many as k-Max centroids. Original *k-Max* = number of nouns to be grouped. Each Firefly is a solution.
4: While (iteration <=t)
5: For Each Firefly The location of each original centroid is the same as the location of the data
6: For Each Centroid on Fireflies
7: Measure the distance to another centroid in the same Firefly
8: Centroids with a distance less than (*p x furthest_distance*) will be combined by selecting the new centroid as the new position
9: For all data
10: Measure the data distance with all centroids
11: Cluster the data to the nearest centroid (using the feature similarity equation)
12: Count the quality of each batch
13: Move the Fireflies to the Fireflies with the highest i value.
The movement of the Fireflies towards the Fireflies with the highest i is determined using the X formula
14: iteration++
15: Repeat the steps in lines 4 – 14 as long as the conditions in line 4 are met
16: end while
17: Choose one Firefly that has the highest intensity (i)
18: For each group that is on the selected Fireflies
19: If cluster member (data) is only 1, make it a label/node and stop
20: If there is more than 1 group member
21: Select one data as a label/node (Caraballo method), and divide the rest of the data into two groups (equal division algorithm) and group each member into two new groups.
22: repeat step 19

Algorithm 1. Pseudocode of Bisection-Firefly Algorithm

The Firefly Algorithm has several essential features for performing acquisition tasks (taxonomic learning). The basic properties of the Firefly Algorithm are still preserved to demonstrate that some of the basic principles of the Firefly Algorithm can be used in taxonomic learning. Line 8 of BFA demonstrates how grouping is started. Centroids with a distance less than (*p* x *furthest distance*) will be merged, with the new centroid assigned the new position. Assume there are 50 Fireflies, for example (population). Each firefly has 200 dimensions (D). Each dimension can be a cluster centre (centroid) with as much data as the number of verbs extracted (accumulated) from the text. For every 200 dimensions (D), the distance is measured with nouns (D) that exist in the group, whether they are close and need

to be joined. The *p-value* used as a threshold value is 5% of the furthest distance to form D_{final}. Used 50 populations, each consisting of 200 (D) original dimensions raised randomly. Each dimension, representing a cluster centre (centroid), has as much data as the number of available nouns, as shown in Fig. 1.

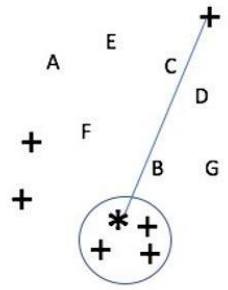

Fig. 1. The sample of the grouping process.

Figure 1 shows an example of grouping with 7 data: A, B, D, E, F, and G. The + sign is the centre of the cluster (centroid), and if the distance between two or more cluster centres is less than the value p x of the furthest distance (straight line), then join and determine the median value (* sign). If the value of *Dfinal* is 5, then the total number of groups that will be formed is only 5. For each data to be clustered, the algorithm will find the smallest d value in each population. The slightest error value means that the cluster formed is a cluster of better quality. This new position is determined by finding the midpoint between all the centroids to be merged. This process is repeated for each blink until several clusters are formed in each blink.

It should be noted that each Firefly has a set of solutions to problems in BFA. Thus, each Firefly initially has a set of centroids whose number is determined based on the amount of data (*k-Max*) extracted from the text. Lines 9 to 11 show how the grouping process was implemented by referring to Fig. 1. For each data (i.e., noun) in each blink, measure distance (equal) to all centroids in the same Firefly only. Data are grouped at the nearest centroid. Then in line 12, the quality of each cluster is measured using an objective function.

This study used the objective function to measure the error by using the square root equation of the sum of the errors. The group with the slightest error has the brightest light intensity (i). Now each blink has a value of *i*. The population (Fireflies) with the smallest objective function value will be the centre or goal of the movement for other Fireflies. Lines 4 to 14 will repeat if the conditions on line 4 are met. Lines 17 to 20 show the process of building the hierarchy. Line 17 states that Fireflies (populations) containing the best-quality clusters will be selected. The Fireflies selected are the Fireflies that have the brightest light intensity level (i), which is the smallest number of errors. The fireflies contain several clusters. Each cluster includes data (nouns). In each group, one of the data will be selected as a label or node based on the algorithm [24], after which the instructions on lines 18 to 21 will be executed for the rest of the members. Lines 18 to 21 are based on the bisector. Each Firefly that represents a group will be divided into two groups. For each group, the result of the division process (in line 18) will be divided again into two. This process will be repeated for each group except the group with only one (1) member.

This division process is called the bisecting process, which splits the selected cluster into two clusters and replaces the original cluster. The cluster bisection is carried out using the K-means basis, with the sum of *k* (the number of new clusters) being two (2). Cimiano [2] has proven that the Bisection K-means is an excellent and fast bisector algorithm. The advantage of using the bisection algorithm is when the initial process of group division or separation is based on "global information" about the objects that are to be grouped because the hierarchical agglomerate grouping method performs the process of building a hierarchy without considering "global information" [25]. Therefore, according to [25], the bisection algorithm produces more accurate hierarchies than the agglomerate algorithm in some applications.

All the proposed clustering methods are developed using Python programming language, and experiments are carried out on workstations with Intel Xeon Processor 3.3 GHz and 4 GB RAM. They are compared in terms of lexical and taxonomic overlap (TO), described in detail in Nazri et al. [26] and other applications [27].

3. Result and discussion

The first comparison that will be discussed is the performance comparison of BKMA and BFA, which is the hybridization of FA and Bisection Algorithm. BKMA and BFA were tested using the same data set. Table 3 and Fig. 2 display the grouping results and comparison between BKMA and BFA.

Table 3. Comparison results between BKMA and BFA.

	Dataset	LR(%)	LP(%)	LF(%)	P_{TO}(%)	R_{TO}(%)	F_{TO}(%)
BKMA	Biochemistry	12.29	35.92	14.82	72	15.25	14.73
	IT	47.83	44.31	35.23	60	17.33	14.04
	Islamic Jur.	32.24	39.92	34.22	65	16.65	14.86
BFA	Biochemistry	14.52	40.43	24.65	90	15.25	30.22
	IT	65.11	52.45	64.43	100	17.33	32.43
	Islamic Jur.	45.54	50.14	52.24	85	16.65	27.01

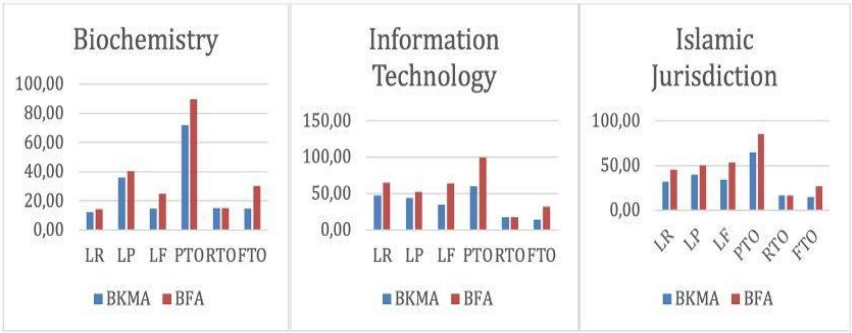

Fig. 2. Comparison results between BKMA and BFA in the visualization of the graph.

Table 3 and Fig. 2 clearly show that the performance of BFA is better than the BKMA in terms of FTO on all datasets. Therefore, it is essential to test whether the results obtained did not occur by chance. The basic idea of this test is to prove that

BFA is better than other cluster methods. The three datasets used to suffer from a high data rarity problem because more than 60% of the terms (nouns) in the data set only have one feature. The feature obtained from the text using the syntactic dependency method clearly shows the severe data rarity problem. This study seeks to develop a robust algorithm when faced with this issue.

4. Conclusions

The major goal of this project is to construct a taxonomy learning algorithm using Indonesian literature in order to enhance the existing hybridization techniques. The Bisection-Firefly Algorithm (BFA) and the Bisection K-Means Algorithm (BKMA) are two hybrid algorithms whose effectiveness is examined in this study (BFA). This empirical study gathers data from tests on three materials written in Indonesian that are related to biochemistry, computer technology, and Islamic law. A comparison of accuracy using the F-measure while addressing data sparsity issues shows that the BFA produces more accurate taxonomies and more resilient and successful results than an alternative method.

References

1. Ismail, A.; Joy, M.S.; Sinclair, J.E.; and Hamzah, M.I. (2009). A metametadata taxonomy to support semantic searching algorithms in metadata repository. *Proceedings of the* 2009 *International Conference on Electrical Engineering and Informatics*, Bangi, Malaysia, 1-6.
2. Cimiano, P. (2006). Ontology learning and population from text: Algorithms, evaluation and applications. (6th ed.). Springer US, 9-17.
3. Ristoski, P.; Faralli, S.; Ponzetto, S.P.; and Paulheim, H. (2017). Large-scale taxonomy induction using entity and word embeddings. *Proceedings of the International Conference on Web Intelligence*, Leipzig, Germany, 81-87.
4. Luu, A.T.; Tay, Y.; Hui, S.C.; and Ng, S. K. (2016). Learning term embeddings for taxonomic relation identification using dynamic weighting neural network. *Proceedings of the* 2016 *Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, 403-413.
5. Wang, C.; He, X.; and Zhou, A. (2017). A short survey on taxonomy learning from text corpora: Issues, resources and recent advances. *Proceedings of the* 2017 *Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, 1190-1203.
6. Dickinson, M.; and Ragheb, M. (2011). Dependency annotation of coordination for learner language. *Proceedings of the International Conference on Dependency Linguistics* (*Depling* 2011), Barcelona, Spain, 135-144.
7. Nazri, M.Z.A.; Shamsudin, S.M.; Bakar, A.A.; and Abd Ghani, T. (2008). Using linguistic patterns in FCA-based approach for automatic acquisition of taxonomies from Malay text. *Proceedings of the* 2008 *International Symposium on Information Technology*, Kuala Lumpur, Malaysia, 1-7.
8. Harris, Z.S. (2002). The background of transformational and metalanguage analysis. In Nevin, B.E. (Ed.), *The Legacy of Zellig Harris : language and information into the* 21*st century*. *Volume* 1, *Philosophy of science, syntax, and semantics*. Amsterdam ; Philadelphia : J. Benjamins Pub. Co.

9. Firth, J.P. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*. Longman,10-32.
10. Hadj Taieb, M.A.; Zesch, T.; and Ben Aouicha, M. (2020). A survey of semantic relatedness evaluation datasets and procedures. *Artificial Intelligence Review*, 53(6), 4407-4448.
11. Ellis, N.C. (2019). Essentials of a theory of language cognition. *The Modern Language Journal*, 103(Suppl 1), 39-60.
12. Casillas, M.; Brown, P.; and Levinson, S.C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, 91(5), 1819-1835.
13. Wang, J.; and Dong, Y. (2020). Measurement of text similarity: A survey. *Information*, 11(9), 421.
14. Berner, J.; Grohs, P.; and Jentzen, A. (2020). Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *SIAM Journal on Mathematics of Data Science*, 2(3), 631-657.
15. Bi, H.; Sun, J.; and Xu, Z. (2018). A graph-based semisupervised deep learning model for PolSAR image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4), 2116-2132.
16. Qiu, J.; Qi, L.; Wang, J.; and Zhang, G. (2018). A hybrid-based method for Chinese domain lightweight ontology construction. *International Journal of Machine Learning and Cybernetics*, 9(9), 1519-1531.
17. Cheng, L.; Yaghoubi, V.; Van Paepegem, W.; and Kersemans, M. (2021). Mahalanobis classification system (MCS) integrated with binary particle swarm optimization for robust quality classification of complex metallic turbine blades. *Mechanical Systems and Signal Processing*, 146, 107060.
18. Moodi, F.; and Saadatfar, H. (2022). An improved K-means algorithm for big data. *IET Software*, 16(1), 48-59.
19. Sun, D.; Fei, H.; and Li, Q. (2018). A bisecting K-Medoids clustering algorithm based on cloud model. *IFAC-PapersOnLine*, 51(11), 308-315.
20. Lewis, S.M.; and Cratsley, C.K. (2008). Flash signal evolution, mate choice, and predation in fireflies. *Annual Review of Entomology*, 53, 293-321.
21. Jaikla, S.; Lewis, S.M.; Thancharoen, A.; and Pinkaew, N. (2020). Distribution, abundance, and habitat characteristics of the congregating firefly, Pteroptyx Olivier (Coleoptera: Lampyridae) in Thailand. *Journal of Asia-Pacific Biodiversity*, 13(3), 358-366.
22. Fister, I.; Fister Jr, I.; Yang, X.-S.; and Brest, J. (2013). A comprehensive review of firefly algorithms. *Swarm and Evolutionary Computation*, 13, 34-46.
23. Senthilnath, J.; Omkar, S.N.; and Mani, V. (2011). Clustering using firefly algorithm: performance study. *Swarm and Evolutionary Computation*, 1(3), 164-171.
24. Caraballo, S.A. (1999). Automatic construction of a hypernym-labeled noun hierarchy from text. *Proceedings of the* 37*th Annual Meeting of the Association for Computational Linguistics*, College Park, Maryland, USA, 120-126.

25. Desikan, K.; and Grace, G.H. (2013). Optimal clustering scheme for repeated bisection partitional algorithm. *International Journal of Engineering Research and Application*, 3(5), 1492-1495.
26. Nazri, M.Z.A.; Shamsuddin, S.M.; Bakar, A.A.; and Abdullah, S. (2011). A hybrid approach for learning concept hierarchy from Malay text using artificial immune network. *Natural Computing*, 10(1), 275-304.
27. Zakaria, M.N.A.; Ahmed, A.N.; Malek, M.A.; Birima, A.H.; Khan, M.M.H.; Sherif, M.; and Elshafie, A. (2023). Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia. *Heliyon*, 9(7), e17689.