Semi-Automatic Concept Map Generation Approach of Web-Based Kit-Build Concept Map Authoring Tool

Aryo Pinandito (✉)

Hiroshima University, Higashihiroshima, Japan
Universitas Brawijaya, Malang, Indonesia
`aryo@ub.ac.id`

Didik Dwi Prasetya

Hiroshima University, Higashihiroshima, Japan
Universitas Negeri Malang, Malang, Indonesia

Yusuke Hayashi, Tsukasa Hirashima
Hiroshima University, Higashihiroshima, Japan

Abstract—Apart from contributing to students' learning outcomes, learning activities with digital concept maps were useful, fun, and engaging. Kit-Build concept map is a learning framework that incorporated concept map recomposition as its essential activity. Students learn through recomposing digital concept maps from a set of teacher's concept map components; hence, a teacher concept map is essential in Kit-Build. In composing a teacher concept map of Kit-Build, teachers should reflect the learning context and strategy, embody their purpose and intention, students' understanding level, and focus questions in the forms of concept maps. Automatic generation approach produces general concept maps that are perceived to be unsuitable in corresponding said reasons; thus, a semi-automatic approach becomes preferred. The Kit-Build concept mapping tool has been extended with a support function that semi-automatically generates concept maps with Concept Map Mining approach. The design of the extraction and summarization phase of the concept map generation process, which suggests the extracted concepts and proposition triples to the teachers, is presented in this study. However, the accuracy of the support system's suggestions has yet to be discovered before the tool being considered to be used in a real learning environment of EFL reading comprehension with Kit-Build and therefore investigated in this study. The result suggested that the proposed Kit-Build concept map authoring support tool is better used to refine a concept map in more detail.

Keywords—Accuracy, concept map, Concept Map Mining, EFL, Kit-Build

1 Introduction

A concept map can be identified as a form of boundary objects. A boundary object is a tool, artifact, or scaffold to mediate discussion and negotiation between two or

more different views [1] and could be realized in any form of tangible objects. Elaborating concept maps into learning activities helped students depict and explore their understanding [2], and improve their learning achievement [3-5]. Their interaction performance could improve if the activities were supported by a computer-supported concept mapping tool [6]. Thus, learning and assessment through interactive activities [7] could be conducted with more fun and engaging [8,9].

Computer and mobile devices have become an enabler to greater access to learning contents in distance learning. In a situation where learning activities have to be conducted remotely, the need for mobile learning becomes more imminent [10]. Several studies in learning strategy have incorporated concept maps and computer-based concept mapping as one interactive learning activity [7,11] during learning. Another study also involved artificial intelligence in processing students' concept maps to assess and predict their understanding [12]. Further development of a learning strategy with digital concept maps brings up a learning framework called Kit-Build concept map to quickly and easily assess students' understanding [13].

Learning effects have been confirmed in learning activities that use Kit-Build. Students learn through concept map re-composition of a Kit-Build concept map kit, a set of concept map components of a teacher concept map. Kit-Build has been found useful in many trials, classrooms, and subjects, such as university-level math [14], university-level computer science [15], geography in junior high school [16], and science in elementary school [17].

Kit-Build concept map is also being used to support learning English as a Foreign Language (EFL) reading comprehension [18, 19]. Students use their computer tablets and mobile devices to access the learning contents and recompose concept maps to represent their understanding of the readings. The use of mobile devices in learning EFL could help students improve their understanding more [20]. With the provided concept map analysis tool, Kit-Build concept map framework helped the instructor quickly gain insight into the development of students' knowledge through the recomposition and comparison analysis. In an EFL learning strategy that uses Kit-Build concept map, many readings were used. Thus, many concept maps, which represented the readings, also have to be recomposed by the students for practices. In contrast to composing a traditional open-end concept map, students compose their concept maps by recomposing a concept map kit; hence, the teacher concept map is essential in Kit-Build [17]. Moreover, recomposing a concept map from components could help students focus more on concepts and ideas represented by the components [21].

The main drawback in learning with Kit-Build is that teachers have to prepare a concept map before using the map to explain the learning subject or be decomposed into a kit for the students to recompose. Composing a good concept map of an English text was difficult and time-consuming [19,22]. Preparing the concept maps has been an obstacle for teachers to adopt Kit-Build as their teaching strategy. Moreover, the concept maps have to adequately represent the text while corresponding to the teacher's strategy.

Concept maps could be automatically generated with various techniques and approaches [23-26]. The generation process could incorporate the Concept Map Mining (CMM) method that involved the Natural Language Processing (NLP) and text min-

ing. Even though concept maps could be generated automatically from texts, the generated concept maps have issues regarding coverage, accuracy, readability, and suitability [27] and yet satisfying to be practically used in supporting learning EFL reading comprehension with Kit-Build. Therefore, semi-automatically generating the concept maps with human interference becomes an option to address the issue.

This study presented the design of the generation approach of a concept map authoring tool that adopted the CMM approach and text mining to generate concept maps of EFL reading comprehension texts semi-automatically. The text processing of the authoring support tool was implemented and developed using web technology, and therefore the application was generally accessible through web and mobile devices. The designed authoring support tool was semi-automatically generating the concept maps by suggesting keywords and proposition triples for the teachers to choose, modify, and incorporate the suggestions into their concept maps. However, before the tool is considered to be used practically with Kit-Build, the performance of the authoring support system has to be evaluated. Therefore, the accuracy of the suggestions yielded by the system, and how the tool is perceived, need to be discovered.

Assessing an education application could depict how the application is used and discover how it further affected the learning [28, 29]. However, before evaluating the learning outcomes, several experts were involved. They used and evaluated the performance of the authoring support tool in suggesting concept map components of EFL reading comprehension texts and assisting their concept map composition activity. In order to guide the study, the following research questions were addressed:

1. With the designed extraction approach towards EFL reading comprehension learning strategy with Kit-Build, what is the accuracy of the suggested keywords and proposition triples of the authoring support tool?
2. Will the developed concept map authoring support tool perceived to be useful for assisting teachers in composing concept maps from EFL reading comprehension texts and support their teaching activities with Kit-Build concept map?

The results suggested that the authoring support tool provided good accuracy in suggesting the concepts; the proposition triple suggestions were acceptable. The tool was perceived as useful to assist teachers compose concept maps of EFL reading comprehension texts, especially in refining a concept map more detail. The design of semi-automatic generation approach is presented in the Literature Review; the methodology and the EFL learning strategy with Kit-Build concept map are presented in Methodology section; the results are discussed in the Result and Discussion section; the remaining sections conclude the results and present the limitation and future work.

2 Literature Review

2.1 Kit-build concept map

Kit-Build concept map is a learning framework that incorporated high-directed concept maps in its learning strategy. Concept map re-composition is the key activity

in learning with Kit-Build concept map. Students recomposed a set of concept map components—called a kit—to represent their understanding of a particular learning topic. Using concept maps in digital form to compose and recompose concept maps, Kit-Build offered a quick and easy assessment of students' understanding by comparing students' concept maps with a respective teacher concept map [17,30]. The comparison and analysis could be easily performed with the provided authoring tool and analyzer. The general process of learning with Kit-Build is depicted in Fig. 1.
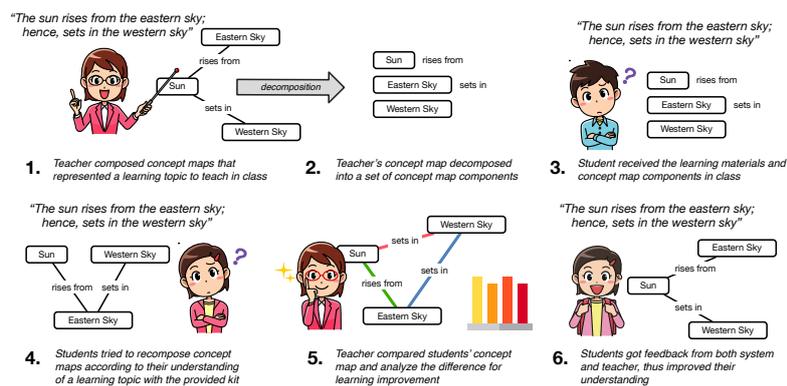

Fig. 1. General activities of learning with Kit-Build concept map

As previously mentioned, students recomposed their concept map from a Kit-Build kit that is a decomposition of a teacher's respective concept map. Before students recomposed concept maps, teachers have to compose concept maps and decompose the map into a kit before conducting the actual learning activity with Kit-Build concept map. Teachers have to carefully plan and compose their concept maps beforehand because their concept maps held important concepts and ideas of a learning topic. Thus, the concept map in which be decomposed into a kit could help the students focused on important ideas represented by the kit. Extending the concept mapping activity [15] and incorporating Kit-Build into collaborative learning [31,32] could also improve their interaction during learning [21,33] and help students to understand and comprehend the learning material.

2.2 Concept's extraction and prioritization

Concept's labels in a concept map are commonly specified from a document keywords as they most likely represent the document's topics or main ideas. There are many ways to determine keywords for a text document, but most existing approaches use manual assignments by experts or defined manually based on the concept map's authors' judgment. However, with the NLP and text mining techniques, it is possible to extract keywords from a text document based on statistical approaches.

Most concept labels in a concept map are nouns or noun phrases. Therefore, it is evident that one approach to extracting concepts from a text is by capturing all of the

nouns and noun-phrases tagged from NLP's annotation process. However, not all nouns or noun-phrases are essential to consider to put into a concept map. Subject to using the concept maps, a concept label should be non-trivial and essential.

The relevance of words or keywords to a document can be measured by their frequency of appearance [34]. The term Term Frequency-Inverse Document Frequency (TF-IDF) is a popular algorithm in calculating the weight of terms or keywords to a document. The TF-IDF weight (w_{ij}) of a term i in a document j can be computed from the number of occurrences of the term in the document (tf_{ij}), the number of documents containing the term (df_i), and the total number of documents (N) in a corpus as in (1).

$$w_{ij} = tf_{ij} \times \log \frac{N}{df_i} \tag{1}$$

The similarity between two keywords can also be computed by finding the cosine similarity ($cos(\theta)$) between vectors of two keywords ($\vec{d_j}$, $\vec{d_k}$) composed of n unique words from both vectors. The similarity value could be calculated based on the TF-IDF weight of the composing words in each keyword (w_{ij}, w_{ik}) as in (2).

$$cos(\theta) = \frac{\vec{d_j} \cdot \vec{d_k}}{|\vec{d_j}| \cdot |\vec{d_k}|} = \frac{\sum_{i=1}^{n} w_{ij} w_{ik}}{\sqrt{\sum_{i=1}^{n} w_{ij}^2} \sqrt{\sum_{i=1}^{n} w_{ik}^2}} \tag{2}$$

An unsupervised method for extracting keywords from a text document, namely Rapid Automatic Keyword Extraction (RAKE), measures the importance of a keyword to the document based on its composing words' degree and frequency [35]. The degree of a word is defined as the frequency of words appear in the keyword candidate list plus its frequency of co-occurrence with other words in the candidate list. The score of a keyword (S_R), which composed of n words, can be calculated as a sum of the ratio of each composing word's degree ($deg(w_i)$) and its frequency of appearance ($freq(w_i)$) in the candidate list as in (3). Important keywords can be selected from top T scoring keywords from the list or by setting a minimum keyword score. When working with keywords on a single document, RAKE is also more computationally efficient than a graph-based ranking approach, such as TextRank [35].

$$S_R = \sum_{i=1}^{n} \frac{deg(w_i)}{freq(w_i)} \tag{3}$$

2.3 Relationship extraction

In the extraction process of propositions from a text, a CMM system should identify concepts and relationships (links). Identifying an accurate and meaningful relationship between a link and two concepts plays an essential role in forming a concept map proposition. A concept map proposition can be represented as a triple of concept-link-concept or a set of subject-relation-object in open information extraction (Open IE). The Open IE annotation process of Stanford CoreNLP annotates and extracts triples from a text, representing a subject, relation, and relation object [36]. The Open IE annotation corresponds to the open domain relation that captures the relation phrases expressed by the combination of verb-nouns patterns [37] and a natural logic classifier

[36]. The triples resulting from the syntactic relationship extraction process serve potential candidates for concept map propositions.

Relationships in a complete sentence were mostly constructed by verbs. However, many sentences use a combination of verb-noun phrases instead of a single verb to depict a meaningful relationship between subjects and objects. Therefore, identifying relationship by relying on a single verb is insufficient to identify one good and meaningful relationship. A simple regular expression pattern can be applied to a sentence's POS tags annotation sequence to identify verbs and verb-noun phrases as a potential relationship candidate for propositions in a concept map. The extraction process that uses the pattern to extract relationships from a sentence is called the syntactic relationship extraction [36]. The pattern is given by (4)

$$V \mid VP \mid VW * P \qquad\qquad (4)$$

where
V   =   verb particle? adverb?
W   =   noun | adjective | adverb | pronoun | determiner
P   =   preposition | particle | information marker

Furthermore, the pattern could reduce the number of uninformative relationships extracted from a sentence [37] and improve the extraction performance [38].

The computed distance between two keyword vectors represents the similarity level between keywords. The similarity among two keywords or phrases using cosine similarity measure can be determined by (a) calculating the keywords' TF-IDF value with (1), (b) transforming the keywords into vectors space model, and (c) computing the distance of both vectors with cosine similarity measure as in (2).

3   Methodology

This study focused on supporting the concept map composing activity of teacher concept map semi-automatically. Support was given in the form of recommendations of concepts and propositions extracted from English texts. The extraction process adopted the CMM approach that involved NLP and text mining techniques. In addition to presenting the design of the extraction process, the accuracy of the yielded recommendations was analyzed and evaluated.

The Kit-Build concept map authoring tool was designed to be used on tablet computers and implemented in HTML5 and Javascript technology. The support feature extended the current authoring tool and could also be run on the same platforms. Users could use the tool with the new support feature with their personal computers or tablet devices. This study extended the current Kit-Build concept map authoring tool's functionality by adding a recommendation system as an authoring support feature and enriched the way teachers composed and improved their concept map. The support function recommended keywords and proposition triples while also allowed modifications made to the suggested items.

The design and development of the authoring support function are shown in Fig. 2. Before the design and development of the support function were conducted, a prelim-

inary study regarding Kit-Build concept map framework, CMM, and the Stanford CoreNLP toolkit was carried out to identify and analyze how CMM and the NLP toolkit were able to extract concepts and propositions from English reading texts. How teachers get assistance from the support function was designed following the strategy of learning EFL reading comprehension with Kit-Build concept map.
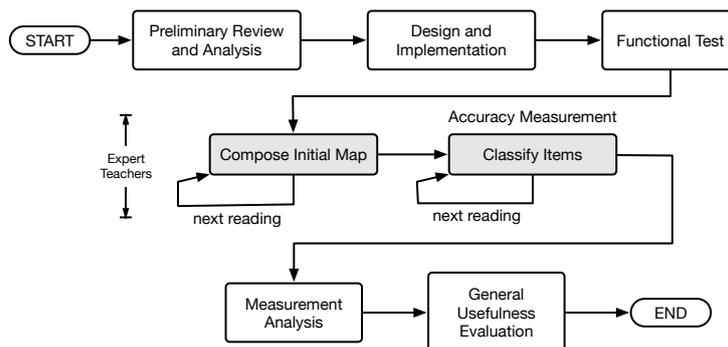

Fig. 2. Research Methodology

Review and analysis of the current Kit-Build concept map tool were carried out to identify how the authoring support function could be integrated into the current Kit-Build concept map tool, thus resulting in a general requirements specification. The support system architecture, activities, and text processing were designed according to the requirements specification. After all of the designs were implemented into the target program and prototype, several tests were carried out to ensure the system work as designed, thus yielded the expected outcomes.

The support system performance regarding support function accuracy was evaluated using several English reading comprehension materials. Several English teachers evaluated the system and composed concept maps using the support function of the authoring tool. The yielded suggestions of the support function were evaluated and classified by the teachers per their initial concept maps to measure the accuracy of suggested items. Additionally, the evaluators were given a questionnaire and requested to evaluate the tool's support function regarding their perceived usefulness in contrast to the traditional concept mapping approach with Kit-Build concept map.

3.1 EFL reading comprehension learning strategy

The concept mapping strategy that was applied to this study consisted of several steps. First, teachers created a concept map that sophisticatedly represented the reading and decomposed the map into a set of concept map components (Kit-Build kit). The map would be used in the next phase of learning with Kit-Build concept map framework, where students reconstructed the map to express their understanding regarding the reading. One sophisticating concept map should have enough relevant concepts and relationships to represent the content. The map should not be too general to represent the text with merely a small number of concepts and relationships or be

overwhelmed with many complicated concepts and unnecessary relationships that might confuse the students.

During the comparison analysis phase, students' concept maps were compared with a teacher concept map. The comparison identifies their misconceptions or misunderstanding regarding the reading pointed by the different and missing parts. The teachers were then further explaining the reading and refining their concept map to a more detailed concept map that better cover the missing and different parts. In this learning strategy of EFL reading comprehension with Kit-Build concept map, teachers composed their concept map at least twice, i.e., during the initial concept map composing activity and during the refinement of their initial concept map. Therefore, the authoring support features were expected to support their composing activity in these two situations.

3.2   Support function evaluation

To evaluate the support system's performance, three English teachers were selected as expert evaluators based on their expertise in using concept maps in teaching EFL reading comprehension. They were requested to compose their initial concept maps of a sophisticated level based on Mueller's concept map rubric by using the system with support function [39]. Upon reviewing several publicly available concept map evaluation rubrics, they had agreed that Mueller's concept map rubric was compatible with the strategy of learning EFL reading comprehension with Kit-Build concept map; hence, used in this research.

Fifteen reading comprehension texts of Barron's TOEFL iBT learning materials [40] were selected to evaluate the accuracy performance of support function. During their initial concept map composing, teachers were requested to use the authoring tool's support features to get recommendations of keywords and proposition triples. They were given a tutorial and practicing to use the concept map authoring tool.

A recommendation system's accuracy can be evaluated from the positive predictive value (PPV) and the true-positive rate (TPR). In information retrieval, PPV and TPR are generally called by precision and recall, respectively. PPV considers both true-positive items (t_p) and false-positive items (f_p), while TPR considers both true-positive items and false-negative items (f_n). PPV or precision was calculated with (5), while TPR or recall was calculated with (6).

$$\text{PPV} = \frac{t_p}{t_p + f_p} \tag{5}$$

$$\text{TPV} = \frac{t_p}{t_p + f_n} \tag{6}$$

In evaluating the suggested keywords and proposition triples' accuracy, the teachers were asked to identify and classify every keyword and triple from the suggestion list for its relevance and appropriateness to their concept maps. The support function accuracy performance of this study was measured by F-measure (F_1). F-measure is commonly used to evaluate the performance of information retrieval systems such as search engines, machine learning models, and natural language processing. Both pre-

cision and recall values were considered in calculating the system's accuracy performance. The F-measure is formalized in (7).

$$F_1 = 2 \cdot \frac{\text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} = \frac{t_p}{t_p + \frac{1}{2}(f_p + f_n)} \tag{7}$$

True-positive items (t_p) were defined as the number of suggested items and were used as part of their concept maps, either with or without slight modifications. The modifications may include adding or renaming part of the concept's or link's label; or altering the suggested triples. On the contrary, false-positive items (f_p) were defined as the number of suggested items not used in their concept maps. Part of their concept maps, which were not suggested by the support tool and had to be created manually, were identified as false-negative items (f_n). True-negative items were irrelevant in this study, because the system did not suggest the items or used in the concept map.

In this study, the support system was used in two concept mapping activities. First, it was used during the initial concept map composing; and second, it was used during the refinement of teachers' concept map. As the purpose of both concept mapping activities was different, the accuracy of the suggestions might differ. Thus, suggested concepts and propositions, which were not included in teachers' initial concept maps but potentially used to refine the concept maps, were classified differently. These potential suggestions were classified as false-positive potential (f_{pp}). Therefore, false-positive potential items were counted as false-positive and true-positive for the initial and the refinement concept mapping, respectively. The precision value of the support function for the initial concept mapping was calculated with (8), and the precision value for the refinement concept mapping activity was calculated with (9).

$$\text{PPV} = \frac{t_p}{t_p + (f_{pp} + f_{p)}} \tag{8}$$

$$\text{PPV} = \frac{t_p + f_{pp}}{(t_p + f_{pp}) + f_p} \tag{9}$$

4 Result and Discussion

4.1 Architecture design and support processing flow

The system was built on top of web technologies, thus accessible from most computers and mobile devices with access to the Internet. A web server served the application interface and provided the data Application Programming Interface (API) to communicate with the database. Both the NLP toolkit and Web API transferred the data in JavaScript Object Notation (JSON) format through the Hypertext Transfer Protocol (HTTP). The processing layer of the support system is depicted in Fig. 3.

The support system processing logic was integrated into the Kit-Build concept map web application. The essential part of the processing layers was the NLP toolkit and served by the Stanford CoreNLP server. The toolkit pre-processed, parsed, and annotated the text input. The Stanford CoreNLP annotated the English input text with

tokenization, sentence splitting, lemmatization, part-of-speech (POS) tagging, dependency parsing, coreference resolution, and open information extraction. The POS tagging annotation identified and labeled noun and noun-phrases for keyword identification processing. Each identified noun or noun-phrase was calculated for its weight by the TF-IDF method. Similarities among keywords can be computed with the cosine similarity measure. RAKE ranked the identified keywords according to the text.
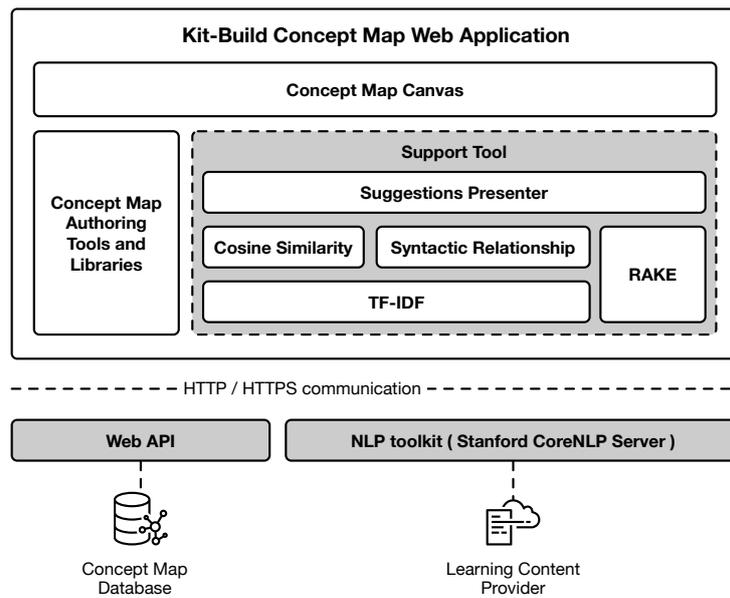

Fig. 3. Processing architecture layer of concept map support system.

The Stanford CoreNLP also yielded Open IE triples that already in a subject, relation, relation-object format. All triples, whose subject or relation-object has a strong similarity to the extracted keywords would be suggested to the teachers. However, teachers can also opt to see all of the Open IE annotated triples. The Syntactic Relationship Extraction process extracted possible relationships and triples based on the regular expression patterns as in (4). According to the review of the annotated text, some of the Open IE annotation results were unsatisfactory. Several triples resulting from the Open IE annotation were considered unnecessary, and verb-based linking words cannot be properly identified. Thus, the Syntactic Relationship Extraction (SRE) was incorporated to identify verb-based relationships of a sentence.

The system's suggestion processing flow in recommending the keywords and proposition triples and the relevant application layers that process the document text is shown in Fig. 4. The processing flow was initiated when teachers opened a text file. The processing flow was divided into two parts to optimize the processing time. The first part—the pre-processing—ran immediately after the NLP toolkit had annotated the text. The process extracted essential keywords from the text and calculated the

similarity among keywords. The second part ran during sentence navigation and processed the currently selected sentence.
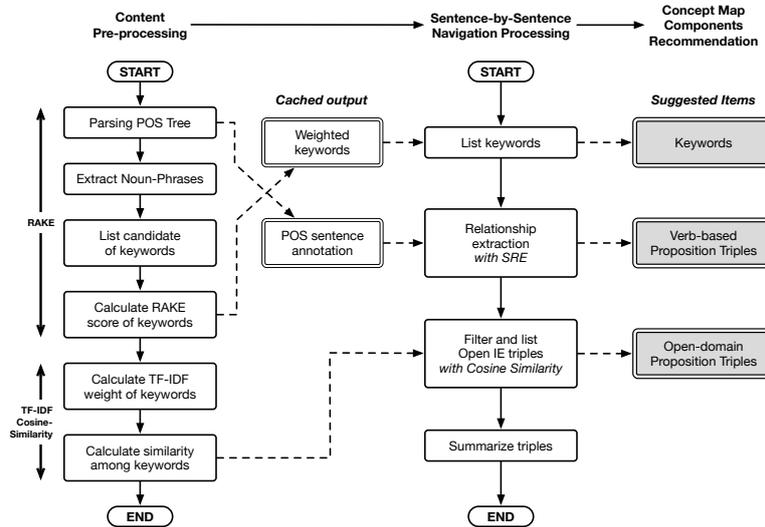

Fig. 4. Recommendation processing flow.

4.2 Implementation result

The support system's interface and functionality were mostly implemented in web programming languages. The Web API was implemented in the PHP programming language. The cosine similarity, RAKE, SRE, and TF-IDF processors were implemented with Javascript and ran client-side. The graph visualization of the concept map used the Cytoscape.js library [41] with several plugins added.

The keywords and proposition triples were served on a sentence-by-sentence basis, which means that the given suggestions reflected a sentence selected by the teacher. The suggestions were presented to the teacher in a support dialog, as depicted in Fig. 3. Using the support dialog, teachers can navigate the text to get suggestions for another sentence. Teachers could modify the triples and put the desired keywords into the canvas from the suggested keywords and proposition triples. They can also modify the propositions, organize the concept map layout, and alter the map's visual appearance with the existing simple concept map composing tool.

As depicted in the support dialog in Fig. 5, the suggestions were divided into three sections. Each section represented the suggestions from RAKE, SRE, and Open IE. The topmost part, which consisted of three suggested proposition triples, was obtained from the SRE method. The center part, which consisted of one suggested proposition triple, was obtained from the filtered Open IE method. Lastly, three keywords were suggested from the RAKE approach. The label of a proposition triple's subject and relation object can be modified or replaced with any suggested keywords. Thus, forming a more appropriate proposition before incorporated the triples as part of the con-

cept map. Keyword added to the canvas would represent a concept in the concept map. Additionally, teachers could compose a proposition triple manually.
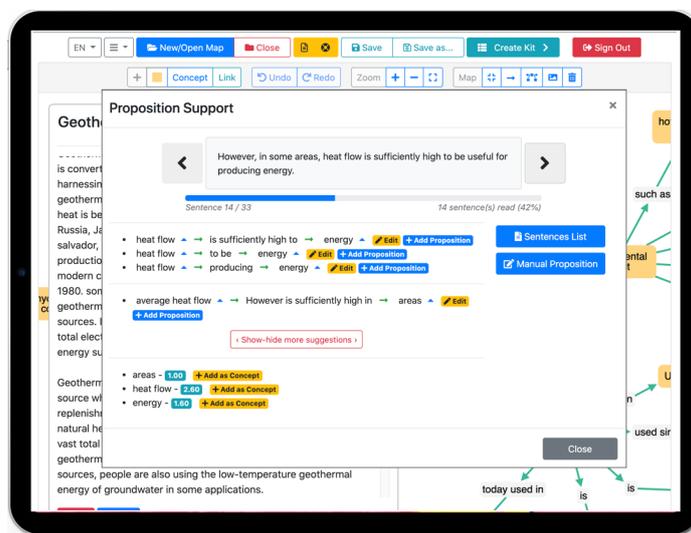

Fig. 5. The Kit-Build concept map authoring tool; running and displaying the authoring support dialog on a computer tablet device.

4.3 Initial concept mapping using the authoring support tool

The system evaluation involved 15 EFL reading texts. According to the evaluation result, on average, the support tool suggested 157 keywords, 85 proposition triples of the SRE approach, and 62 proposition triples of the optimized Open IE. The SRE method extracted and recommended verb-based proposition triples according to the POS pattern of a sentence [37], while Open IE triples corresponded to open domain-based relations [36]. If all said keywords and propositions were visualized into a concept map, a relatively huge concept map would be generated, and evaluating such a concept map would be challenging. As generating proper and satisfying concept maps was difficult to achieve, the semi-automatic approach proposed was presumed a better approach than a fully-automated or manual approach, at least in this study.

On average, the EFL readings used to evaluate the support features were composed of 719 words and 34 sentences with a standard deviation of 50 words and six sentences, respectively. The description of the EFL reading comprehension text used in the evaluation of the tool, the number of sentences (ns), the number of suggested keywords (nk), the number of suggested propositions from the syntactic relationship extraction (nps), and the number of suggested propositions from the filtered open IE triples (npo) are shown in Table 1.

Table 1. Number of suggested keywords and proposition triples of the readings

ID	Title	ns	nk	nps	npo
T1	Beowulf	39	159	82	57
T2	Building with Arches	34	151	88	58
T3	Civilization	24	150	81	48
T4	Exotic and Endangered Species	45	168	83	63
T5	Four Stages of Planetary Development	38	141	89	49
T6	Geothermal Energy	33	148	70	58
T7	Looking at Teathre History	29	155	97	26
T8	Organic Architecture	34	168	90	76
T9	Paleolithic Art	29	141	84	49
T10	Piaget's Cognitive Development Theory	41	139	119	49
T11	Resources and Industrialism in Canada	23	175	64	107
T12	Rising Sea Levels	38	195	80	85
T13	The Evolution of Birds	35	144	84	55
T14	The Hydrologic Cycle	38	156	71	68
T15	Thermoregulation	33	169	87	82

During the initial concept mapping activities, teachers created 15 sophisticating concept maps—one concept map for each reading, resulting in concept maps composed of 35 concepts and 36 propositions on average. Composing concept maps with such numbers of concepts and propositions was not an easy task. Moreover, teachers were requested to compose many concept maps in a relatively short time. The descriptive statistics of the expert teacher's initial concept maps during authoring support system evaluation are shown in Table 2.

Table 2. Descriptive statistics of expert teacher's initial concept maps

Expert Teacher (ET)	n	Average		Minimum		Maximum		Standard Deviation	
		Concepts	*Props.*	*Concepts*	*Props.*	*Concepts*	*Props.*	*Concepts*	*Props.*
ET#1	15	29.47	29.87	20	19	61	61	9.912	10.449
ET#2	15	43.87	43.33	25	24	89	90	14.773	15.230
ET#3	15	31.27	35.67	17	24	43	58	7.698	9.5196
Overall		34.9	36.3	17	19	89	90	12.896	13.204

4.4 Support function accuracy performance measurement

Three expert teachers classified the suggestions into a statistical confusion matrix, representing instances in the predicted classes and instances in the actual classes based on their initial concept maps. Each suggestion is classified into one of true-positive, false-positive, false-negative, or true-negative class. The classifications process was also carried out to all groups, i.e., RAKE keywords, SRE propositions triples, and filtered Open IE triples.

The number of relevant (true-positive (t_p)), potentially relevant (false-positive potential (f_{pp})), irrelevant suggestions (false-positive (f_p)), and non-existing suggestions (false-negative (f_n)) of one of the experts are shown in Table 3. The classifications

were carried out to the suggested keywords, proposition triples of SRE, and proposition triples of Open IE. These values will be used to calculate the support system accuracy from their precision and recall value.

Table 3. Classification of Suggested Items of a teacher

ID	Keywords				Verb-based Triples				Open-domain Triples			
	t_p	f_{pp}	f_p	f_n	t_p	f_{pp}	f_p	f_n	t_p	f_{pp}	f_p	f_n
T1	130	17	12	3	41	17	22	17	31	11	12	29
T2	67	63	20	2	28	29	28	6	25	11	22	14
T3	58	84	8	3	26	41	9	22	12	17	18	24
T4	68	86	16	4	46	18	21	6	17	31	13	8
T5	46	82	13	1	34	27	27	10	17	18	16	10
T6	68	68	11	1	29	19	20	16	17	20	19	17
T7	54	82	19	5	17	50	24	13	7	18	0	16
T8	43	111	15	2	20	30	34	19	12	37	24	22
T9	64	62	14	4	32	28	27	2	13	20	14	5
T10	43	82	14	4	22	65	31	17	15	27	7	26
T11	53	116	4	0	17	40	7	11	44	14	43	10
T12	70	112	15	3	49	7	20	24	52	2	25	23
T13	53	86	4	6	21	47	16	16	18	26	11	18
T14	63	89	6	4	18	31	16	25	15	26	19	28
T15	74	92	3	3	32	41	13	9	24	26	31	10

The summary of the support system performance and the standard deviation value of the concept mapping support tool are shown in Table 4. The table depicts the system's average precision, recall, and F-measure values for 15 EFL reading comprehension texts of three raters. For the suggested keywords, the support system has a relatively low precision but high recall value. However, triples obtained from the SRE and Open IE approach have low precision and recall value. A low precision value indicated a high number of false-positives; similarly, a low recall value indicated a high number of false-negative items, depicting few suggestions were used in the initial concept maps. A low recall value indicated that many parts of their concept maps have to be created manually. It can also be said that the system could not suggest satisfying proposition triples for use in the teachers' initial concept maps.

The support system's accuracy performance can be depicted from the F-measure value, which is the harmony between precision and recall values. According to Table 3, it can be seen that the system has the lowest F-measure average performance of 42%, 14.9%, and 19.2% in terms of keywords, open domain-based triples, and verb-based triples, respectively. The highest F-measure average performance is 55.9%, 40.3%, 44.8% for the keywords, open domain-based triples, and verb-based triples. Supposing the same suggestions were used to refine the concept map, the F-measure performance improved for a maximum of 95.1%, 69.4%, 76.7% for the keywords, open domain-based triples, and verb-based triples, respectively.

**Table 4.** Support Tool Accuracy Performance

Suggestion Type	n	Precision		Recall		F-measure	
		mean	*s.d.*	*mean*	*s.d.*	*mean*	*s.d.*
Expert Teacher #1							
Potential suggestions (f_{pp}) are counted as false-positives (f_p)							
Keyword	15	0.405	0.129	0.953	0.0283	0.559	0.112
Open domain-based triple	15	0.333	0.121	0.541	0.169	0.403	0.119
Verb-based triple	15	0.347	0.131	0.667	0.148	0.448	0.132
Potential suggestions (f_{pp}) are counted as true-positives (t_p)							
Keyword	15	0.926	0.0359	0.98	0.0113	0.951	0.02
Open domain-based triple	15	0.694	0.112	0.71	0.107	0.693	0.063
Verb-based triple	15	0.73	0.085	0.815	0.0866	0.767	0.068
Expert Teacher #2							
Potential suggestions (f_{pp}) are counted as false-positives (f_p)							
Keyword	15	0.294	0.108	0.925	0.0368	0.436	0.116
Open domain-based triple	15	0.13	0.115	0.184	0.127	0.149	0.121
Verb-based triple	15	0.145	0.0904	0.311	0.109	0.192	0.102
Potential suggestions (f_{pp}) are counted as true-positives (t_p)							
Keyword	15	0.931	0.0385	0.973	0.02	0.951	0.025
Open domain-based triple	15	0.519	0.122	0.497	0.138	0.502	0.12
Verb-based triple	15	0.668	0.074	0.696	0.0774	0.676	0.049
Expert Teacher #3							
Potential suggestions (f_{pp}) are counted as false-positives (f_p)							
Keyword	15	0.278	0.0842	0.916	0.0458	0.42	0.102
Open domain-based triple	15	0.207	0.107	0.331	0.123	0.243	0.097
Verb-based triple	15	0.219	0.0486	0.535	0.17	0.306	0.069
Potential suggestions (f_{pp}) are counted as true-positives (t_p)							
Keyword	15	0.896	0.0424	0.972	0.0163	0.932	0.024
Open domain-based triple	15	0.758	0.0817	0.649	0.113	0.694	0.082
Verb-based triple	15	0.749	0.076	0.784	0.106	0.761	0.07

When teachers created their initial concept maps, all false-positive potential items were regarded as false-positives items, resulting in a larger number of false-positive items, thus lowering the system's precision value. In composing the initial concept maps, they created one sophisticated concept map rather than directly composed a very detailed and huge concept map. Thus, it was expected that a lower number of true-positive items was noticed in their initial concept maps. However, in another use of the support function where they refine their concept map, false-positive potential items were potentially useful to refine the concept map and could be regarded as true-positive items; hence, improved precision.

4.5 Usefulness evaluation

A questionnaire based on the Technology Acceptance Model (TAM) questionnaire was given to the expert teachers to get their usefulness impression regarding the support tool. The questionnaire consisted of 5 questions on a 7-Likert scale, which has a Cronbach alpha value of 0.84, indicating good reliability. The questionnaire was giv-

en in comparison with the manual concept mapping approach. Their responses to the questionnaire are shown in Fig. 6.
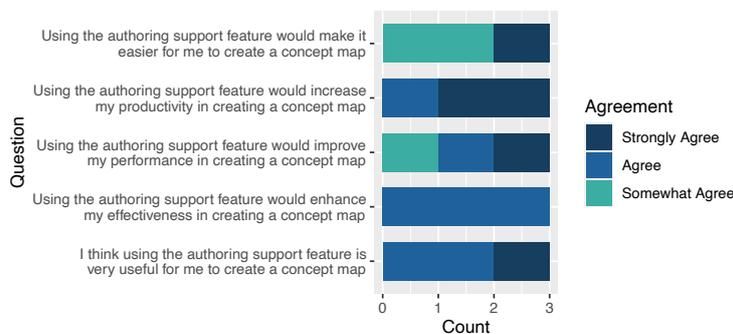
Fig. 6. Expert teacher's perceived usefulness of using Kit-Build concept map authoring tool with support function

According to the questionnaire result, as shown in Fig. 6, none of them gave neutral or negative feedback regarding the usefulness. They agreed that the authoring support tool would enhance the effectiveness and increase their productivity in creating concept maps from text learning materials; thus, they thought the support tool was very useful. However, two of them somewhat agreed that the authoring support feature would make them easier in composing concept maps because the way they composed their concept maps with the authoring support tool was more complex than a manual composition. Their responses regarding the support tool's usefulness were in line with the tool's usability level, as presented in [22]. Hence, improvement in usability could improve the usability of the authoring support tool. Nevertheless, it can be said the developed Kit-Build concept map authoring tool was useful in assisting teachers in composing concept maps from English texts.

5   Limitation and Future Work

The result of the proposed semi-automatic concept map generation showed a promising result to be used in the actual EFL learning environment with Kit-Build concept map. Thus, practical use and trial of the authoring support tool with EFL teachers for use in EFL reading comprehension learning will be the future work of this study.

The concept map generation approach does not elaborate on teachers' logic and thinking while composing their concept maps. Following the teachers' logic in determining which concepts and relationship triples are important and genuinely relevant to the topic is important could improve the suggestions' accuracy. Discovering factors that influence how teachers compose their concept maps is also one important aspect of suggesting more accurate extraction and further improving the suggestions. Specifying the context of learning and adapting the automatic generation approach to con-

form to the learning context could potentially yield better relevance keywords and propositions; hence, more accurate and better components to suggest.

Technically, the extraction approach applied does not elaborate coreference and semantic resolution in which they may improve how the system resolves pronouns and identify similar words in a text. Elaborating coreference resolution could improve the extraction approach. Incorporating semantic approach and additional processing to support ill-structured English sentences are several other potential future works of this study towards a better concept map generation process. Even though the extraction approach in this study was not specifically designed and limited for EFL reading comprehension use, it is also possible to use the tool to extract concept map components from another English text. However, it might yield inaccurate and unexpected results.

6　Conclusion

Assessing an education application could depict the application performance when it is used and also discover the drawbacks of the application. While this study does not evaluate how the concept map authoring tool could impact the learning, the prototype of the tool itself was used and evaluated by experts to depict its performance. The experts, who used Kit-Build in their teaching, also reviewed the tool regarding the usefulness of the support tool in assisting concept map composition of EFL reading comprehension texts.

The tool's functionality has been extended with an authoring support function in addition to the standard authoring features of the existing Kit-Build concept map tool. The support function semi-automatically generates concept maps with the CMM approach in assisting teachers in composing their concept maps. The assistance was given semi-automatically in the form of keyword and proposition triple recommendations. In the experiment, the recommendations were extracted from EFL reading comprehension learning material using a combination of CMM, NLP, and text mining methods. Three experts were involved in using the support function and evaluated its performance. They perceived that the developed concept map authoring tool is useful; thus, ascertaining the tool could help teachers compose concept maps from English learning material. However, some limitations in using the tool exist, and further improvement regarding the tool's usability was needed.

Even though the resulting concept maps were subject to the teacher's subjectivity, the proposed semi-automatic concept map generation approach provided by the authoring support tool could achieve an average-to-good accuracy level. Concerning the learning strategy of EFL reading comprehension of this research, accuracy level of 55.9%, 40.3%, and 44.8% could be attained from the respective RAKE keywords, open domain-based triples, and verb-based triples of the initial concept maps. Better accuracy could be achieved if potential suggestions were incorporated to refine the concept maps, yielding an improved accuracy of 95.1%, 69.3%, and 76.7% for the respective RAKE keywords, open domain-based triples, and verb-based triples. According to the accuracy measurement results, it can be concluded that the tool support

feature was performed exceptionally well, especially in supporting teachers to refine and improve their concept maps through a semi-automatic concept map generation approach.

7 Acknowledgement

8 References

[1] F. Saeidifard, K. Heidari, M. Foroughi, and A. Soltani, (2014), "Concept mapping as a method to teach an evidence-based educated medical topic: A comparative study in medical students," Journal of Diabetes and Metabolic Disorders, vol. 13(1), pp. 86-90, https://doi.org/10.1186/s40200-014-0086-1

[2] S. Y. Wu, S. Y. Chen, and H. T. Hou, (2016), "Exploring the interactive pat- terns of concept map-based online discussion: A sequential analysis of users' operations, cognitive processing and knowledge construction," Interactive Learning Environments, vol. 24(8), pp. 1778-1794, https://doi.org/10.1080/10494820.2015.1057740

[3] S. Tseng, (2020), "Using concept mapping activities to enhance students' critical thinking skills at a high school in Taiwan," Asia-Pacific Edu Res 29, pp. 249-256, https://doi.org/10.1007/s40299-019-00474-0

[4] C. T. Machado and A. A. Carvalho, (2020), "Concept Mapping: Benefits and Challenges in Higher Education," The Journal of Continuing Higher Education, vol. 1-16, pp. 38-53, https://doi.org/10.1080/07377363.2020.1712579

[5] D. Astriani, A. R. Purnomo, H. Susilo, H. Suwono, and B. Lukiati, (2020), "Mind Mapping in Learning Models: A Tool to Improve Student Metacognitive Skills", International Journal of Emerging Technologies in Learning (iJET) Vol. 11(11), pp. 4-17, https://doi.org/10.3991/ijet.v15i06.12657

[6] U. Joshi and S. Vyas, (2018), "Assessment of perception and effectiveness of concept mapping in learning epidemiology," Indian Journal of Community Medicine, vol. 43(1), pp. 37-39, https://doi.org/10.4103/ijcm.ijcm_375_16

[7] Z. Ma and L. Shi, (2016), "Application of Visual Interactive Concept Map in College English Writing Teaching", International Journal of Emerging Technologies in Learning (iJET) Vol. 11(11), pp. 32-36, https://doi.org/10.3991/ijet.v11i11.6248

[8] C. H. Lai, B. S. Jong, Y. T. Hsia, and T. W. Lin, (2020), "Applying Concept Map to Game-Base Concept Assessment", International Journal of Interactive Mobile Technologies (iJIM) Vol. 14(6). pp. 159-170. https://doi.org/10.3991/ijim.v14i06. 12057

[9] S. Papadakis, (2020). "Robots and Robotics Kits for Early Childhood and First School Age", International Journal of Interactive Mobile Technologies (iJIM), Vol. 14(18), pp. 34–56, https://doi.org/10.3991/ijim.v14i18.16631

[10] M. Drolia, E. Sifaki, S. Papadakis, and M. Kalogiannakis, (2020). "An Overview of Mobile Learning for Refugee Students: Juxtaposing Refugee Needs with Mobile Applications' Characteristics," Challenges Vol. 11(2), 31. https://doi.org/10.3390/challe11020031

[11] B. A. Leauby, K. A. Szabat, and J. D. Maas, (2010), "Concept mapping-an empirical study in introductory financial accounting", Accounting Education, vol. 19(3), pp. 279–300, https://doi.org/10.1080/09639280903412334

[12] B. J. Daley, A. J. Cañas, and T. Stark-Schweitzer, (2007), "CmapTools: Integrating teaching, learning and evaluation in online courses", New Directions for Adult and Continuing Education, vol. 113, pp. 37–47, https://doi.org/10.1002/ace.245

[13] D. B. Hay, (2007), "Using concept mapping to measure deep, surface and non-learning outcomes", Studies in Higher Education, vol. 32(1), pp. 39–57, https://doi.org/10.1080/03075070601099432

[14] L. Sadita, T. Hirashima, Y. Hayashi, W. Wunnasri, J. Pailai, K. Junus, and H. B. Santoso, (2020), "Collaborative concept mapping with reciprocal kit-build: a practical use in linear algebra course", Research and Practice in Technology Enhanced Learning, 15(1), https://doi.org/10.1186/s41039-020-00136-6

[15] D. D. Prasetya, T. Hirashima, Y. Hayashi, (2020), "Study on Extended Scratch-Build Concept Map to Enhance Students' Understanding and Promote Quality of Knowledge Structure", The International Journal of Advanced Computer Science and Applications (IJACSA) Vol. 11(4), pp: 144-153, https://doi.org/10. 14569/ijacsa.2020.0110420

[16] Y. Hayashi, T. Nomura, and Hirashima, T, (2020), "Prediction of Group Learning Results from an Aggregation of Individual Understanding with Kit-Build Concept Map", presented in the International Conference on Artificial Intelligence in Education, Springer, Cham., pp. 109-113.https://doi.org/10.1007/978-3-030-52240-7_20

[17] J. Pailai, W. Wunnasri, K. Yoshida, Y. Hayashi, and T. Hirashima, (2017), "The practical use of Kit-Build concept map on formative assessment", Research and Practice in Technology Enhanced Learning, vol. 12(20), https://doi.org/10.1186/s41039-017-0060-x

[18] M. Alkhateeb, T. Rajab, Y. Hayashi, and T. Hirashima, (2016), "Experimental Use of Kit-Build Concept Map System to Support Reading Comprehension of EFL in Comparing with Selective Underlining Strategy", International Journal of Advanced Computer Science and Applications, 7(4), pp. 80–87, https://doi.org/10.14569/ijacsa.2016.070410

[19] B. S. Andoko, Y. Hayashi, T. Hirashima, and A. N. Asri, (2020), "Improving English reading for EFL readers with reviewing kit-build concept map", Research and Practice in Technology Enhanced Learning, vol. 15(7), pp. 1–19. https://doi.org/10.1186/s41039-020-00126-8

[20] R. Ivanova, A. Ivanov, and Z. Nikonova. (2020), "Application of Mobile Technologies in Foreign Language Learners' Project Activity, " International Journal of Interactive Mobile Technologies (iJIM) Vol. 14(21) pp. 64-77, https://doi.org/10.3991/ijim.v14i21. 18471

[21] A. Pinandito, Y. Hayashi, and T. Hirashima, (2020), Online Collaborative Kit-Build Concept Map. Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education Vol I, pp. 189-191.

[22] A. Pinandito, H. M. Az-zahra, T. Hirashima, and Y. Hayashi, (2019), "User Experience Evaluation on Computer-Supported Concept Map Authoring Tool of Kit-Build Concept Map Framework", presented at 2019 International Conference on Sustainable Information Engineering and Technology (SIET), pp. 289–294, https://doi.org/10.1109/siet48054.2019.8986005

[23] K. Zubrinic, D. Kalpic, and M. Milicevic, (2012), "The automatic creation of concept maps from documents written using morphologically rich languages", Expert Systems with Applications, vol. 39, no. 16, pp. 12709–12718, https://doi.org/10.1016/j.eswa.2012.04.065

[24] I. Qasim, J. W. Jeong, J. U. Heu, and D. H. Lee, (2013), "Concept map construction from text documents using affinity propagation", Journal of Information Science, pp. 1–18, https://doi.org/10.1177/0165551513494645

[25] B. N. Wafula, (2016), "Automatic Construction of Concept Maps", Master Thesis. [Online]. Available: http://urn.fi/urn:nbn:fi:uef-20160829

[26] A. Nugumanova, M. Mansurova, E. Alimzhanov, D. Zyryanov, and K. Apayev, (2015), "Automatic Generation of Concept Maps based on Collection of Teaching Materials", presented at the 4th International Conference on Data Management Technologies and Applications, pp. 248–254.https://doi.org/10.5220/0005554702480254

[27] T. Atapattu, K. Falkner, and N. Falkner, (2017). "A comprehensive text analysis of lecture slides to generate concept maps," Computers & Education 115, pp. 96–113. https://doi.org/10.1016/j.compedu.2017.08.001

[28] S. Papadakis, (2020). Tools for evaluating educational apps for young children: a systematic review of the literature. Interactive Technology and Smart Education, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/itse-08-2020-0127

[29] S. Papadakis, J. Vaiopoulou, M. Kalogiannakis, and D. Stamovlasis, (2020). "Developing and Exploring an Evaluation Tool for Educational Apps (E.T.E.A.) Targeting Kindergarten Children". Sustainability, 12, 4201.https://doi.org/10.3390/su12104201

[30] T. Hirashima, K. Yamasaki, H. Fukuda, and H. Funaoi, (2015), "Framework of Kit-Build Concept Map for Automatic Diagnosis and Its Preliminary Use", Research and Practice in Technology Enhanced Learning, vol. 10(1,17), https://doi.org/10.1186/s41039-015-0018-9

[31] W. Wunnasri, J. Pailai, Y. Hayashi, and T. Hirashima, (2018), "Reciprocal kit-build concept map: An approach for encouraging pair discussion to share each other's understanding," IEICE Transactions on Information and Systems, E101D (9), pp. 2356–2367, https://doi.org/10.1587/transinf.2017edp7420

[32] L. Sadita, P. G. F. Furtado, T. Hirashima, and Y. Hayashi, (2020), "Analysis of the similarity of individual knowledge and the comprehension of partner's representation during collaborative concept mapping with reciprocal Kit Build approach", IEICE Transactions on Information and Systems, E103.D(7), pp. 1722–1731, https://doi.org/10.1587/transinf.2019edp7305

[33] L. Sadita, T. Hirashima, Y. Hayashi, P. G. F. Furtado, K. Junus, and H. B. Santoso, (2020), "The effect of differences in group composition on knowledge transfer, group achievement, and learners' affective responses during reciprocal concept mapping with the Kit-Build Approach", Research and Practice in Technology Enhanced Learning Vol. 15(13). https://doi.org/10.1186/s41039-020-00133-9

[34] G. Salton and C. Buckley, (1988), "Term-weighting approaches in automatic text retrieval," Information Processing & Management Vol. 24(5), pp. 513-523, https://doi.org/10.1016/0306-4573(88)90021-0

[35] S. Rose, D. Engel, N. Cramer, and W. Cowley, (2010), "Automatic keyword extraction from individual documents," in Text Mining Applications and Theory, John Wiley & Sons, US, pp. 4-20, https://doi.org/10.1002/9780470689646.ch1

[36] G. Angeli, M. J. J. Premkumar, and C. D. Manning, (2015), "Leveraging Linguistic Structure for Open Domain Information Extraction," presented at 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing Vol. 1, pp. 344-354.https://doi.org/10.3115/v1/p15-1034

[37] A. Fader, S. Soderland, and O. Etzioni, (2011), "Identifying Relations for Open Information Extraction," presented at the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP11), Association for Computational Linguistics, pp. 1535-1545.

[38] R. Patel, Y. Yang, I. Marshall, A. Nenkova, and B. C. Wallace, (2018), "Syntactic Patterns Improve Information Extraction for Medical Search," presented 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Vol. 2, pp. 371-377, https://doi.org/10. 18653/v1/n18-2060

[39] J. Mueller, "Concept map rubric," [Online], Available: http://jfmueller.faculty.noctrl. edu/240/conceptmaprubric.htm

[40] P. J. Sharpe, (2013), "Barron's TOEFL iBT, 14th edition," in Barron's Test Prep, US.

[41] M. Franz, C. T. Lopes, G. Huck, Y. Dong, O. Sumer, and G. D. Bader, (2015), "Cytoscape.js: a graph theory library for visualisation and analysis", Bioinformatics, vol. 32(2), pp. 309–311, https://doi.org/10.1093/bioinformatics/btv557

9   Authors

Aryo Pinandito is a lecturer of Information System Department and a member of Mobile, Game, and Multimedia Research Group at Computer Science Faculty, Universitas Brawijaya, Indonesia. He is currently pursuing his doctoral degree from Learning Engineering Laboratory, Information Engineering Department, Graduate School of Engineering, Hiroshima University, Japan. Recently, he has been granted an international award as the Best Poster Presentation Award of ICCE2020. Email: aryo@ub.ac.id

Didik Dwi Prasetya is with Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Malang, Indonesia. He is currently a doctoral student of Information Engineering Department, Graduate School of Engineering, Hiroshima University, Japan.

Yusuke Hayashi is an associate professor of the Graduate School of Engineering at Hiroshima University since 2012. He has been engaged in research on knowledge modeling, ontological engineering, and learning engineering. He has received international awards as the Best Paper Award of ICCE2006 and the Best Technical Design Paper Award of ICCE2015.

Tsukasa Hirashima has been a professor of the Graduate School, Department of Information Engineering, Hiroshima University since 2004. Learning Engineering is his major research field. He has received international awards as the Outstanding Paper Award of EDMEDIA95, the Best Paper Award of ICCE2001 & 2002, Honorable Mention Award of AIED2009, APSCE Distinguished Researcher Award in 2009, and the ICCE2015 Best Technical Design Paper Award.