March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

Knowledge Management Model for Improving Teaching Efficiency of Special Students Using Text-Mining

Won-Cheol Park[1], Koo-Rack Park[*2], Dong-Hyen Kim[3]
[1]Research Scholar, Department of Computer Engineering, Kongju National University, 31080, Rep. of Korea
[*2]Professor, Department of Computer Science & Engineering, Kongju National University, 31080, Rep. of Korea
[3]Ph. D., Department of Computer Engineering, Kongju National University, 31080, Rep. of Korea
pwcfrog@nate.com[1], ecgrpark@kongju.ac.kr[*2], dhkim977@naver.com[3]

Article Info
Volume 83
Page Number: 4479 - 4488
Publication Issue:
March - April 2020

Abstract

Background/Objectives: As for an education in a special school, one special teacher is in charge of one student in one-on-one basis. Since each and every special student has different personality and characteristic, a special teacher should be aware of various information on diverse special students, but it is difficult to be aware of all special students in reality. Therefore, the purpose of this study is to provide information of students, which is similar to the daily report of a special student that a teacher has entered, to a special teacher by analyzing the data.

Methods/Statistical analysis: In this paper, a system that provides information of diverse special students who are similar to the student whom a special teacher is in charge of and offers the guidance and education data to the special teacher is proposed. Daily student reports that are entered by special teachers get converted into data, and the data gets stored in the database by analyzing the content of the daily student reports through text mining technique and determining the category. The data of the special students which is in the category that matches the category that has been stored gets searched, and it is provided to the special teacher. The special teachers can plan more effective education and guidance by using this information.

Findings: The proposed system searches and outputs special student's data which is similar based on the data which stores the information of diverse special students through daily student reports which special teachers have created, and provides the data to special teachers. Previous studies have focused on the study on the characteristics of special students and various education guidance methods. It is difficult to say that it was a complete solution for new teachers who just got appointed to provide a lecture and give education and guidance right away. Therefore, the proposed system conveniently provides various data of diverse special students to special teachers in real-time, and it can help the teachers to correctly and effectively provide education and guidance by using the data that has been provided.

Improvements/Applications: The proposed system would be a good index for special teachers, and it is expected to help to improve the effect of the education. In future, the study to analyze the behaviors of special students by taking video of behaviors of special students that special teachers couldn't have recognized in real-time and automatically determine the category should be continued.

Keywords: *Special School, Special Education, Special Student, Text-Mining, Knowledge Management System.*

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 26 March 2020*

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

1. Introduction

As for the effects of all educations, the educational effects are expected based on the effort and dedication of teachers who provide the education. As for an education in a special school, one special teacher is in charge of one student in one-on-one basis in providing the education. In order to enhance the effect of the education, each and every special teacher must be aware of entire data of special students who have different characteristics from each other and education guidance data. But in reality, the education curriculums and education environment that are appropriate for the actual classes are not sufficiently prepared yet[1]. In addition, even though practical trainings are provided by institutions that foster special teachers and there are a lot of studies on the education curriculums, they are not sufficient to provide tailored education that is suitable for special students[2]. It is difficult for special teachers to be aware of all education cases of diverse special students and details of education plans in reality. Even in case of same disability, each student has different characteristic, and therefore, it is difficult to be aware of data of all special students. Therefore in this study, daily student reports that are created by special teachers get converted into data, and it gets stored in the database by analyzing and classifying the contents of the daily student reports by using text mining technique and determining the category of the student data that has been entered. A system that provides the search results to special teachers as a list by searching and comparing from the database the data of other students that is in same category as the data category of a special student that has been stored is proposed. By collecting and providing data of diverse special students through the proposed system, special teachers can conveniently handle a lot of information in real-time, hence it is expected to enhance the educational effect of the special education and help to improve the educational environment.

2. Methods

2.1 Text-Mining

The volume of data that exists in modern society is getting increasingly bigger, and the speed of the data flow is very fast as well. In such situation, studies to effectively utilize the large volume data are being performed. Most data is unstructured. An effective method to extract and acquire only useful information that a user needs from the unstructured data in the form of text became necessary[3]. The processing procedure and processing technology of analyzing the text of unstructured large volume data and providing the information that the user needs by extracting useful information from data such as patterns and knowledges are called “text mining”[4]. By using TF-IDF (Term Frequency-Inverse Document Frequency) which many search engines use, the appearance frequency of a certain word from the related data and entire data is checked, and the importance of the word is measured[5,6]. The word frequency in the data can be checked by using the following [formula 1].

$$TF(t,d) = \log\left(f(t,d) + 1\right) \qquad (1)$$

If the id value that represents a word is assumed to be t, and the id value that is allocated to the data is assumed to be d, it shall be checked how many time the word t appears in the data d, and the value shall be stored. Since the tf value continues to get bigger as the data volume gets increased, the method that is called “log scale frequency” which uses the log is used[7]. It is a value that is how many times a certain word commonly appears in the entire data group. Which represents the number of documents that the word appears in. If a certain word t is a common word that appears in all data, it will become difficult to output an accurate result when it is reflected to the search result. In order to solve this problem, IDF

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

value is calculated in order to identify a commonly appeared word that is not a specific word and to lower the weight of IDF value. The IDF value can be calculated through the following [Formula 2].

$$IDF(t, D) = \log\left(\frac{D}{1 + d \in D : t \in d}\right) \quad (2)$$

The reason that 1 is added to the denominator is to prevent the denominator from becoming 0 since the denominator would become 0 if a certain word does not exist in the entire data. The value of TF-IDF can be calculated by multiplying TF and IDF values which are calculated through [Formula 1] and [Formula 2]. Text mining is being applied and used in various areas. It is being utilized in various areas such as summarization of newspaper, thesis, and report, extraction of subject and information that are the representative keyword or topic in the document, question and answer system, and processing of classification and grouping of wide-range of data[8].

2.2 KMS(Knowledge Management System)

Knowledge is one of a concept that is not visible by eye, and it refers to more valuable information that has been created by adding experiences and successes through recognition, analysis, understanding and interpretation of a user[9]. Unlike passive information that is acquired from the outside, the knowledge includes active procedure of processing the data initiatively by the person who uses the information, and making determination such as technology, experience, and intelligence. As the importance of the knowledge got bigger, the knowledge was managed and systemized which led to a knowledge management system, and it is being actively applied to the enterprise management area[10].

Knowledge management system has two characteristics in large which are a knowledge repository and a knowledge map. Firstly, the knowledge repository collects, systematizes, and stores knowledge and information. It refers to a database that offers functions which are required when the organization members search necessary information or perform search activities. The knowledge repository performs the role for the organization without being limited by the time and the space. Secondly, the knowledge map is an index of information and data which are owned by the members of the organization which can be shared. The knowledge management system maps the information and knowledge in a structured way[11]. In order to respond to the environment that is becoming global and the environment that the competition is fierce, an enterprise needs a system for knowledge management. To do so, the establishment of the knowledge management system for systematic management of knowledge is necessary[12,13].

3. Proposed Work

3.1 System Configuration Map

The following [Figure 1] is an overall system configuration of the proposed system.

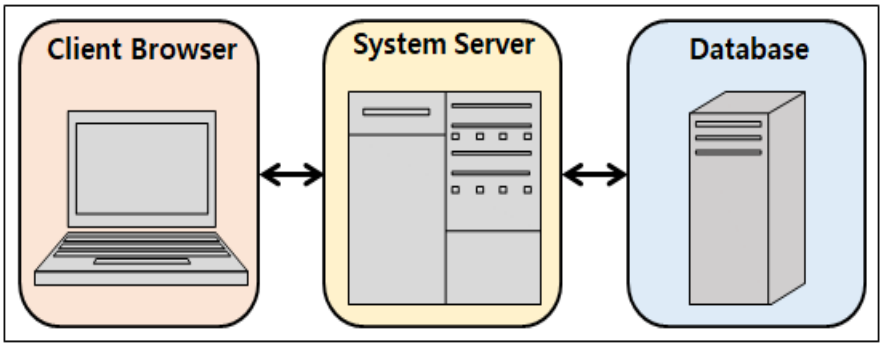

Figure 1. System Configuration.

For analyzing the daily report data of a special student that a special teacher enters by using text mining technique and providing data of special students who are similar to the student that the teacher has entered to special teachers, and it is composed of 3 parts in large.

It is composed of a client browser for special teachers to create and transmit daily reports of special students to the system server, a system server for processing text mining and management of data of the special students that the special teachers have created and uploaded, and a database that the data gets stored. The platform that is used for the proposed system is implemented based on Spring MVC framework in order to enhance the inter-operability by actively using the open source.

3.2 System Process

The following [Figure 2] is a process of the proposed system, and it is composed of 7 steps in large.

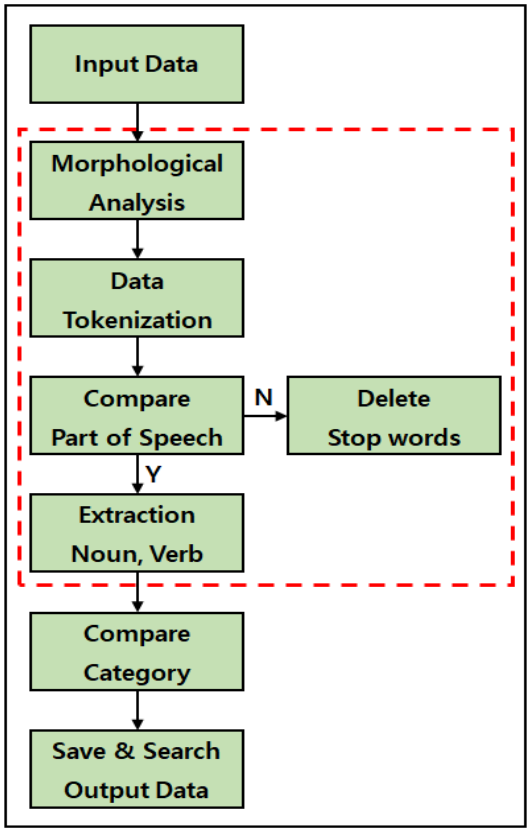

Figure 2. System Process.

Firstly, special teachers enter daily reports or data of special students that they are in charge of into the system. Secondly, it is a step that analyzes and processes the data that has been entered through text mining technique, and the daily report data of special students that has been entered is analyzed in the unit of a morpheme. Thirdly, words that have been analyzed in the unit of a morpheme get tokenized. Fourthly, tokenized words are compared witha part of speech dictionary, and all the unnecessary words except nouns and verbs are removed, and only nouns and verbs are extracted. Fifthly, most highly weighted words among the nouns and verbs that have been extracted get stored temporarily. Sixthly, the highly weighted words that have been selected as the result of the data analysis get compared with the categories that are stored in the previous database for searching a matching category. Once the matching category is searched, it is stored in the database after the category is determined to be the category of current data. Seventhly, data in the same category of the data that has been entered gets searched, and the information of other special students in the same category is provided by displaying the search result to a user as a list.

3.3 Text-Mining

• Morphological Analysis.

The following [Table 1] shows the result that has been acquired in the first step of the text mining which is the morpheme analysis step of special student data that had been entered by special teachers as a table. In this study, the daily student report of a student with intellectual disability was presented as an example data, and the test was performed by entering the data into the system.

After parts of speech of word segments are identified from the data that has been entered, the parts of speech that contain meanings such as nouns, verbs, and adjectives get separated. The number of appearances of each separated word is checked in the data as the appearance frequency.

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

Table 1. Data Stemming Result.

Word	Morpheme(count)	Word	Morpheme(count)
Intellectual	Noun (7)	Indication	Noun (3)
Disability	Noun (7)	Response	Noun (2)
Now	Noun (3)	So	Conjunction (1)
Negatively	Noun (4)	Follow	Verb (4)
Be	Verb (8)	The	Article (14)
Caution	Noun (2)	Have	Verb (5)
Distraction	Noun (2)	Exercise	Noun (2)
Simple	Adjective (1)	Ability	Noun (2)
do	Verb (5)	Good	Adjective (3)

• Delete Stop Word.

Once the morpheme analysis step is completed, the unnecessary parts of speech get filtered for clearer data classification criteria, and then, remaining nouns and verbs are extracted. Each tokenized word gets searched from the parts of speech dictionary that is stored in the database by tokenizing the data that has been entered, and the result that excludes the unnecessary parts of speech gets displayed by comparing them with the parts of speech that had been registered previously.

Firstly, the first phase extraction result that excludes postpositions and endings from the morpheme analysis result data is acquired. Secondly, only nouns and verbs are extracted again from the first extraction result data. The following [Table 2] is the output result of the data that only the nouns and the verbs have been extracted after removing the unnecessary parts.

Table 2. Data Delete Stop Word Result.

Word	Morpheme(count)	Word	Morpheme(count)
Intellectual	Noun (7)	Do	Verb (5)
Disability	Noun (7)	Indication	Noun (3)
Now	Noun (3)	Response	Noun (2)
Negatively	Noun (4)	Follow	Verb (4)
Be	Verb (8)	Have	Verb (5)
Caution	Noun (2)	Exercise	Noun (2)
Distraction	Noun (2)	Ability	Noun (2)

• Category Settings.

In order to determine and classify the category for the special student data that the unnecessary parts have been removed, the following procedures are performed. The result value for the special student data that has been entered gets calculated through the morpheme analysis processing, and then, it gets tokenized by each following word. As for

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

each word that has been tokenized, the appearance frequency for each word is calculated, which shows how many times that each word has appeared in the data. The inverse document frequency gets calculated by taking log on the value that has been acquired by dividing the number of other data that contains tokenized words that the special teacher has entered from the entire data. TF-IDF value can be calculated by multiplying the word frequency and the inverse document frequency. It is searched whether there is a matching category for words with the highest TF-IDF score within the database. Once the matching category is searched, it is determined to be the category of the related data, and the result screen gets displayed in the form of a list by searching the student data in the same category with the related data in the entire database. The following [Figure 3] is a part of the algorithm for outputting the data of the same category after setting the configuration algorithm and the category.

```
Begin
    Read text_data;
    Read all_data_List;
    Read category_List;
    result_1 = Excute Stemming-process(text_data);
    result_2 = Excute tokenize-process(result_1);
    Create wordMap = new wordMap();
    FOR (i ← 0 to result_2.length)
        Insert wordMap ← result_2[i];
        TF = count(result_2[i]) in text_data;
        IDF = log(all_data_List / result_2[i] in wordMap);
        TF-IDF = TF * IDF;
        Insert wordMap.TF-IDF[i] ← TF-IDF;
    End FOR
    category = Find max(wordMap.TF-IDF) in category_List;
    SAVE category(this data);
    adviceData = Find category in max(wordMap.TF-IDF);
    OutPut (Recommendation_Data_List);
End Begin
```

Figure 3. Category setting and result output algorithm.

4. Results and Discussion

The following [Figure 4] is a main screen of the proposed system. The software architecture that is used in implementing the proposed system is as follows. Windows 10 64bit was used as the operating system. JDK 12.0.1 version Java development tool was used and Tomcat 9.0.21 version was used for the web application server. MySQL 8.0.16 version was used for the database. Overall framework was implemented based on SpringMVC framework. As for the main screen configuration, the screen was configured with Word Cloud so that certain keywords with the highest frequency of the use in the distribution graph of target special students which is provided by the Ministry of Education and in the entire database can be easily identified. The lower part of the screen was configured to display the category list and the list of the data that has been recently registered. If an arbitrary data is selected among the data list, the detail page of the related data will get displayed, and the details can be viewed. The upper part of the screen was configured with the buttons that can be used to enter the information and view the information.

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

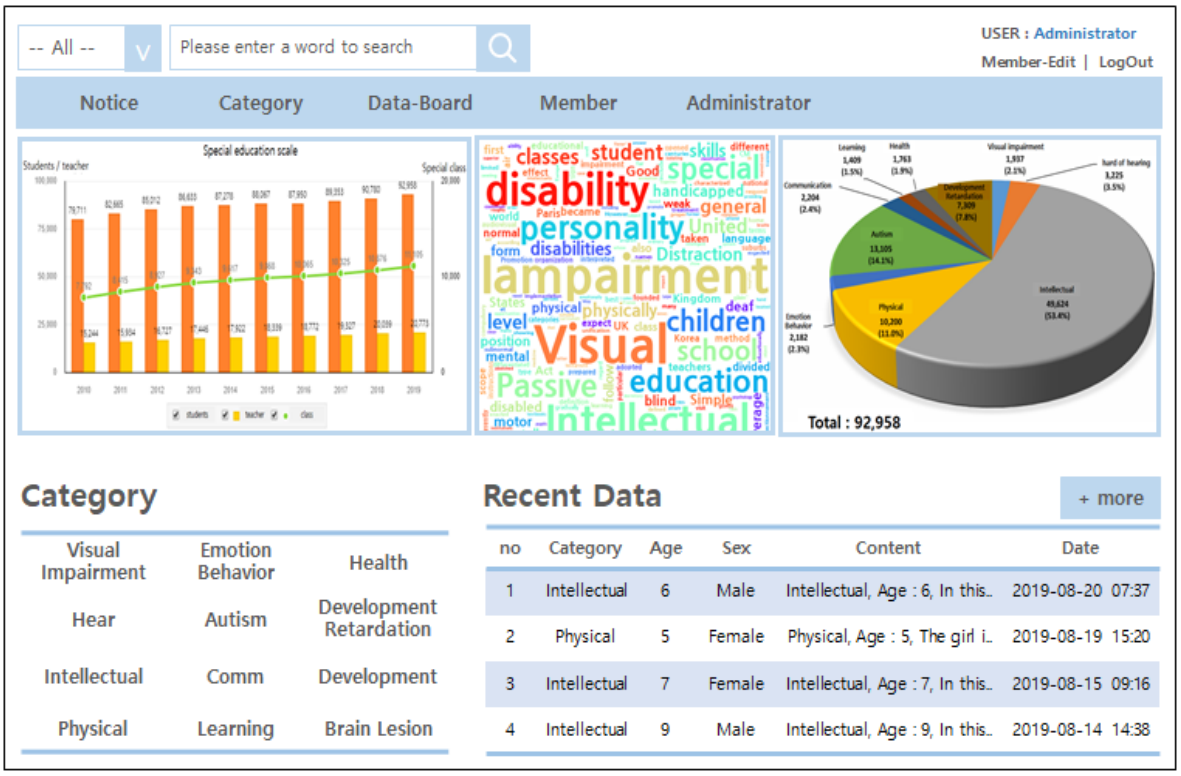

Figure 4. System Main Page.

The following [Figure 5] is the result screen that was displayed by determining a certain keyword to be in a category after processing the data of a special student that has been entered by a special teacher through text data mining, and then visualizing and displaying the words that have been used in the data with Word-Cloud.

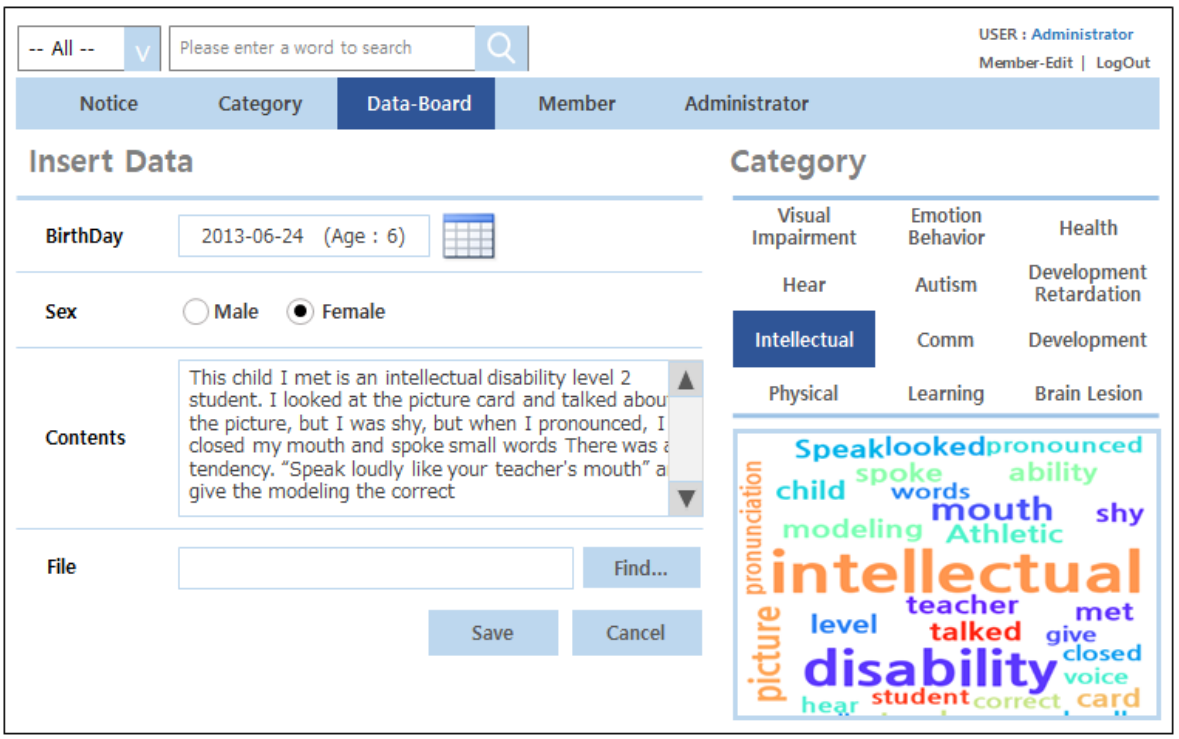

Figure 5. Save & Setting Category.

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

The content of the example data that is used in this study is the content of daily report of a student with intellectual disability. When it is seen from the Word Cloud that is displayed on the upper right side of the screen. It can be checked that the word intellectual disability is clearly displayed in a large font, and it can be checked from the category table on the upper right side of the screen that it is determined to be in the intellectual disability category. The following [Figure 6] is the result that the data in the intellectual disability category which matches the example data that is used in this study is searched in the database and displayed on the screen as a list. It can be checked that the data related to the intellectual disability category which is the same category as the example data that has been used is displayed as the recommended data in the lower part of the screen as a list.

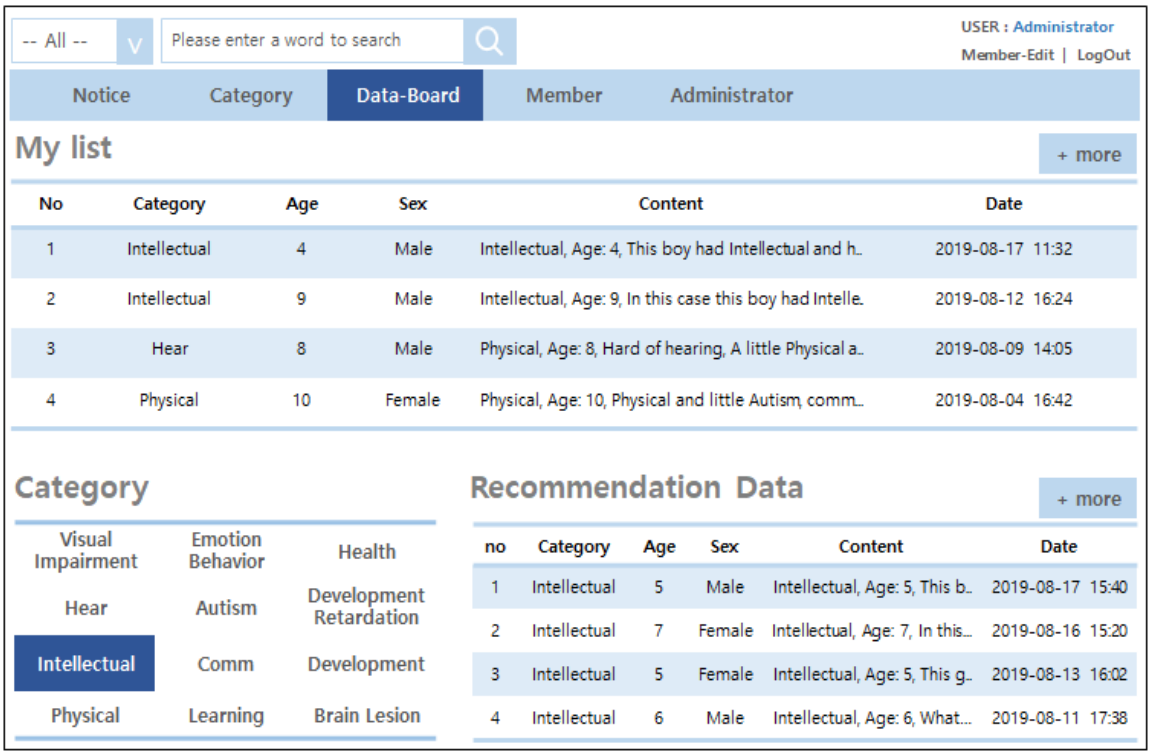

Figure 6. Search & Recommended Data List.

The following [Table 3] shows the result of the accuracy test as a table which was acquired by entering the daily report data of 140 special students who have 9 different disabilities including the example data that has been used in this study. 9 disabilities of special students that have been entered are visual impairment, emotional-behavior disorder, hearing impairment, autistic disorder, intellectual disability, developmental disorder, communication disorder, physical disability, and learning disability. In order to test the accuracy of classifying into categories that the entered data matches the content of the daily report, the test was performed under three evaluation criteria of Accuracy, F-Measure, and Precision among most frequently used methods of measuring the accuracy. The evaluation result for the proposed system was 91.4% on average for the Accuracy, 91.2% on average for F-Measure, and 91.8% on average for the Precision. Since all three evaluation criteria have average score of over 90%, it can be determined that the daily report data that has been entered and the content of the classified category match.

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

Table 3. Proposed System Performance test Result.

Special Student Data (Category)	Accuracy	F-Measure	Precision
Visual Impairment	92.2 %	92.7 %	90.1 %
Emotion Behavior	91.5 %	92.3 %	92.7 %
Hear	92.4 %	89.8 %	90.5 %
Autism	92.7 %	91.5 %	93.2 %
Intellectual	91.4 %	92.1 %	92.4 %
Development	89.7 %	90.6 %	91.7 %
Communication	91.2 %	90.3 %	92.7 %
Physical	90.8 %	91.2 %	90.8 %
Learning	91.4 %	90.6 %	92.1 %

The evaluation result for the proposed system was 91.4% on average for the Accuracy, 91.2% on average for F-Measure, and 91.8% on average for the Precision. Since all three evaluation criteria have average score of over 90%, it can be determined that the daily report data that has been entered and the content of the classified category match.

5. Conclusion

A system that converts daily report records and various cases of special students into data and provides wide range of diverse information to other special teachers in real-time by processing daily reports and information of special students that are ordinarily created by special teachers through text mining technique is proposed. It is the system that helps special teachers to access various information and diverse cases regarding the special students and to establish an appropriate guidance plan accordingly by sorting various cases and wide range of information into categories and providing them to the special teachers. In this study, daily report data of 140 special students who have 9 different disabilities was entered and tested. As the result, the contents that had been entered were correctly classified into the categories that match the contents, and the data that matches the related category was displayed as the recommended data list. If special teachers get to use the proposed system, they would be able to access more diverse and wider range of special students' information, and it is also expected to improve the quality of education since tailored education guidance and diverse education plan can be referred for each individual student by using such information. As for the future study, the research of taking video of certain behaviors of a special student that a special teacher can't observe or notice, converting the record into the data by analyzing the behavior pattern of the special student in the video, and categorizing and storing the data automatically shall be continued.

References

[1] Zigmond N, Kloo A. General and special education are (and should be) different. Handbook of special education. 2011 May;160-172. DOI: 10.4324/9780203837306.ch13.

[2] Anderson N. A., Barksadale M. A., Hite C. E. Preservice teachers' observations of cooperating

March - April 2020
ISSN: 0193-4120 Page No. 4479 - 4488

teachers and peers while participating in an early field experience. Teacher Education Quarterly. 2005;32(4):97-117.

[3] Corredoira R. A., Banerjee P. M. Measuring patent's influence on technological evolution: A study of knowledge spanning and subsequent inventive activity. Research Policy. 2015 Mar;44(2):508-521. DOI: 10.1016/j.respol.2014.10.003.

[4] Tseng Y. H., Lin C. J., Lin Y. I. Text mining techniques for patent analysis. Information processing & management. 2007 Sep;43(5):1216-1247. DOI: 10.1016/j.ipm.2006.11.011.

[5] Soucy P, Mineau G. W. Beyond TFIDF weighting for text categorization in the vector space model. International Joint Conferences on Artificial Intelligence. 2005 Jul;5:1130-1135.

[6] Jiang C, Coenen F, Sanderson R, Zito M. Text classification using graph mining-based feature extraction. Knowledge-Based Systems. 2010 May;23(4):302-308. DOI: 10.1016/j.knosys.2009.11.010.

[7] Trstenjak B, Mikac S, Donko D. KNN with TF-IDF based framework for text categorization. Procedia Engineering. 2014;69:1356-1364. DOI: 10.1016/j.proeng.2014.03.129.

[8] Wang J, Liu J, Wang C. Keyword extraction based on PageRank. In Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2007 May;4426: 857-864. DOI: 10.1007/978-3-540-71701-0_95.

[9] Soualhia A, Maazoun W. E., Affes P. H. Trust and its link with the knowledge management systems. International Journal Economics & Strategic Management of Business Process. 2014 Dec;4(2004):53-60.

[10] Aamodt A, Nygard M. Different roles and mutual dependencies of data, information, and knowledge—an AI perspective on their integration. Data & Knowledge Engineering. 1995 Sep;16(3):191-222. DOI: 10.1016/0169-023X(95)00017-M.

[11] Sambarmuthy V, Subramani M. Special Issue on Information Technology and Knowledge Management. Journal of MIS Quarterly. 2005 Jun;29(2):193-195. DOI: 10.2307/25148676.

[12] Alavi M, Leidner D. E. Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. Journal of MIS Quarterly. 2001 Mar;25(1):107-136. DOI: 10.2307/3250961.

[13] Kankanhalli A, Bernard C. Y., Tan, Wei K. K. Contributing Knowledge to Electronic Knowledge Repositories : An Empirical Investigation. Journal of MIS Quarterly. 2005 Mar;29(1):113-143. DOI: 10.2307/25148670.