Design of Automatic Online Lecture Video Commentator for Visually impaired students supporting Diagram Commentary

Nahyun Eun[1][0000-0001-6119-4468] and JongWoo Lee[2*][0000-0003-1093-7430]

[1] Sookmyung Womens Univ, Dept IT Engn, Cheongpa Ro 47 Gil 100, Seoul 04310, South Korea
dmsskgus@sm.ac.kr
[2] Sookmyung Womens Univ, Dept IT Engn, Cheongpa Ro 47 Gil 100, Seoul 04310, South Korea
bigrain@sm.ac.kr
*Corresponding author

Abstract. With the recent popularization of the Internet, interest in online lectures is increasing. However, due to the characteristics of online lectures that rely on visual materials, visually impaired students need help accessing lecture materials for online lectures. To overcome this, we designed an online lecture system specialized for visually impaired students. Our automatic lecture video content commentator for visually impaired students supporting diagram commentary extracts visual data in the scene of the lecture video, converts it into text, and explains it to students by voice. In particular, a usability evaluation of the commentary technique was conducted to explain the diagram effectively.

Keywords: Online lecture, Visual material, The visually impaired, Video commentary, Diagram Commentary

1 Introduction

Recently, as the internet and mobile devices have become popular and online activities are more closely connected to everyday life, interest in online lectures is also increasing. Online lectures have the advantage of providing customized learning regardless of place and time [1]. Therefore, research is actively underway on how to use this as one of the solutions to increase access to education for all types of people. According to the World Health Organization, visually impaired people have a population of 285 million worldwide [2], meaning that many people have to specialize in services for them when developing online services. Providing specialized technology for the visually impaired is essential because it is sometimes difficult to perform even minor tasks without outside help. A little help can significantly assist their daily lives. Online lectures specialized for them can effectively remove obstacles that visually impaired people take lectures [3]. However, since the online lecture system mainly uses visually shared lecture materials, visually impaired students who accept information with touch and hearing have many difficulties [4]. Therefore, research has been conducted worldwide to identify which part is inconvenient and what the direction of improvement should be [5]. In Korea, a study was conducted to solve this problem by explaining the visual text and the instructor's demonstrative pronouns in the video to the visually impaired [6, 7].

In addition to text, various visual materials in the online lecture system cause the visually impaired inconvenience. To explain formulas in visual data by voice, a study was conducted to accurately convert formulas into sentences and output them to voice by setting the pronunciation rules of formulas expressed in Contents MathML that can express meanings in voice e-books [8]. To explain the graph by voice among visual materials, a GUI program that can convert a mathematical graph into text is used to express the graph in sentences and convert it into a voice [9]. A technique for explaining images using a deep-learning model has also been studied [10]. This technique recognizes images automatically and converts them into sentences, which also makes it easier for visually impaired students to understand images. As such, research on converting visual data into voice has been actively conducted in many ways, but research related to systems that explain “diagrams” into text has not yet been active compared to other fields. The term "diagram" here refers to an image showing the relationship between nodes or objects as an arrow or connection line from one node to another, such as a flowchart, a network diagram, and an ER diagram. These diagrams help students understand better when they first encounter the learning content [11]. Therefore, this study aims to help increase the learning efficiency of the visually impaired by automatically explaining not only text and images but also diagrams by voice.

According to WebAIM, which studies web accessibility for the disabled, mobile usage among the visually impaired increased from 12% to 82% between 2009 and 2014 [12]. In a 2020 survey of people with disabilities by the Korea Institute for Health and Social Affairs, 92.4% of the visually impaired use mobile phones and smartphones [13]. Compared to 34.8% of the visually impaired using a PC, the proportion of the using a mobile is high. In addition, auxiliary technology devices, widely studied to help the visually impaired learn, are less accessible than mobile learning due to availability and economic feasibility [14]. Accordingly, studies have been conducted to improve mobile accessibility for the visually impaired, and it has a positive effect on the lives of the visually impaired [15].

In this paper, we present a method for designing and developing an application that automatically supports voice commentary about the visual material of lecture videos for the visually impaired in a mobile environment. The operation of the application is all progressed through sound and voice. Visual material of lecture videos is classified into text, images, and diagrams and supports voice commentary functions for each. Especially, we focus on generating diagram commentary, which distinguishes between general images and diagrams and generates commentary that considers associations between nodes in the diagram and provides it to visually impaired students by sound. It also helps students learn visual information from the lecture video before the instructor's explanation is played by detecting the time of scene transition in the lecture video, thereby improving their understanding of the lecture content.

2 Related Works

2.1 A mobile e-learning application for enhancement of basic mathematical skills in visually impaired children (2023)

In a mobile e-learning application for enhancement of basic mathematical skill in visually impaired children [16], research was conducted on the development of mobile e-learning applications that help visually impaired children learn math easily. After developing an app using functions of mobile applications available for the visually impaired, such as touch screen design, audio feedback, and feedback using vibration, a usability evaluation was conducted on the visually impaired. As a result of the evaluation, it was found that the application helps improve interest and understanding of mathematics. However, since this app only provides learning about basic mathematical operations, the range of learning content is narrow.

2.2 Voice Commentary System to Increase Understanding of Video Lectures for Visually Impaired Students (2022)

In a [7], a study was conducted on the development of a web program for visually impaired students that explains the visual text in the image and the instructor's pronoun in the lecture video. After conducting a study on the obstacles that visually impaired people experience when taking classes, to solve them, commentary rules for visual text, images, and graphs in the lecture video's screen data are set and read aloud, and explanation of the instructor's emphasis action is provided. As a result of developing a prototype that performs their functions and conducting a usability evaluation, it was found that the case of providing commentary had a higher understanding of the contents of the lecture than the case of not providing it. However, the images described here are simply for conveying visual information, and images for conveying information, such as diagrams, are not interpreted.

2.3 Image Description Guidelines

The DIAGRAM Center, supported by the U.S. Department of Education, researched to increase web accessibility for the visually impaired by proposing a more accessible and cheaper way to explain diagrams through Image Description Guidelines [17]. This paper presents a method for diagram commentary by dividing the diagram into six types, such as picture type, ben diagram, and flowchart, explaining examples of how to explain it by type. Also, it presents a diagram description method that varies the degree of detail of the description by grade of the student who encounters the diagram. This makes it easier for all types of users, including the visually impaired, to access digital content. However, this has the disadvantage that the explanation method accepted may be ambiguous depending on the person because it is not based on accurate rules but only suggests a commentary method through examples.

3 Design of software structure And Methods for interpreting visual materials by Type

3.1 System Overview

The automatic lecture video content commentator for visually impaired students in this paper is a mobile application developed using Flutter [18], an open-source framework developed and supported by Google. It will be available on Android and iOS platforms. The server was developed using the Python [20] language using Flask [19], and the database uses SQLite [21]. Fig. 1 shows the overall system configuration of the automatic lecture video content commentator.

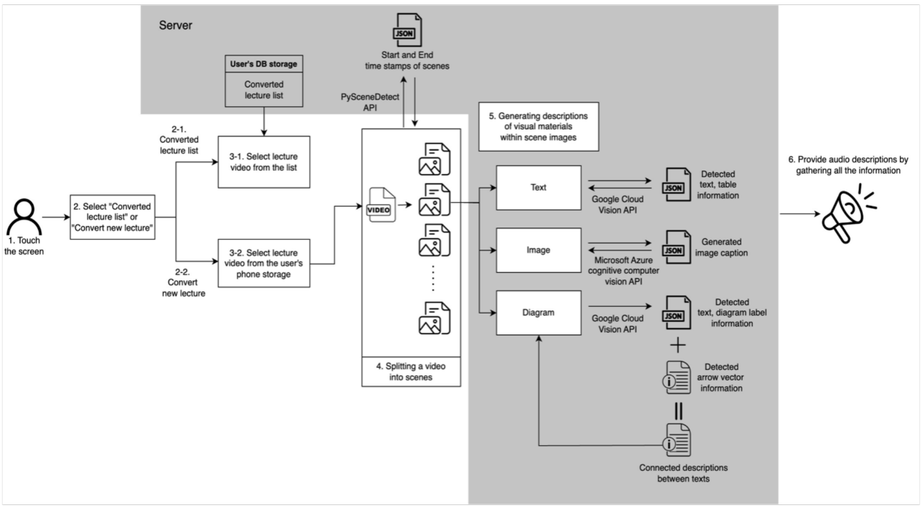

Fig. 1. An overall system diagram of an automatic lecture video content commentator for visually impaired students

First, when a user runs our app and touches the screen, a voice guide is provided to help them access the previously converted lecture list or select from the function of converting a new lecture. Users can answer through voice, and if they want access to the previously converted lecture list, they can access the lecture video list stored in the server's database so that the user can take one of the videos on the list. Lecture videos stored in the database store files for previously converted lecture video content commentary in pairs. If the user wants to convert a new lecture, he or she can select the video from the user's mobile phone's local storage. The selected video is sent to the server for scene-specific frame extraction of the video, and the frames of the video are analyzed using the PySceneDetect API [22]. The PySceneDetect API is a video frame analysis tool written in Python that detects scene changes in video and automatically distributes video to individual clips. Time stamps of the beginning and end points of video shots are output using the API, and images of each shot are stored. This scene image file is used to analyze visual materials such as text, images(pictures), and diagrams. Specific analysis and commentary methods of visual materials will be described in detail later. After the analysis and commentary of all scenes divided by time stamps are completed, all information is generated as a single JSON file and stored in the user's database. At the same time, the contents of the JSON file are converted into sound and printed so that the visually impaired can understand the contents of the lecture material before listening to the lecture video. The playback sequence of information of the converted visual materials and original video for one video can be represented in Fig. 2. If the user wants to listen to the lecture material once more before moving on to another scene, the application can request it.

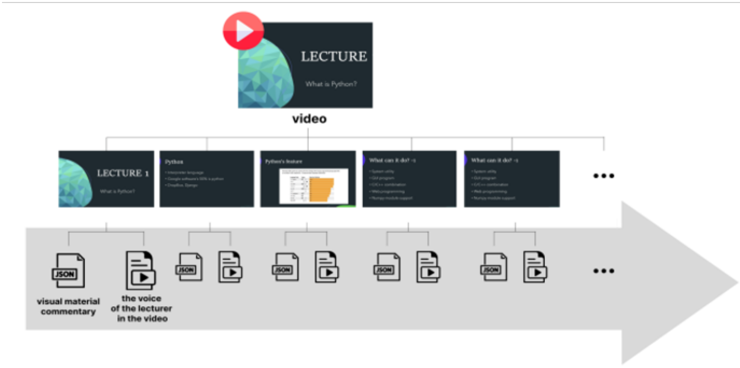

Fig. 2. The order of the converted files for one video

The automatic lecture video content commentator for visually impaired students is a mobile application implemented with Flutter. Since it is an application for the visually impaired, it provides an intuitive and more straightforward UI than general applications. The application operates through sound, listens to the user's voice, and performs commands. Fig.3 shows the overall prototype implementation screen of the application is shown.

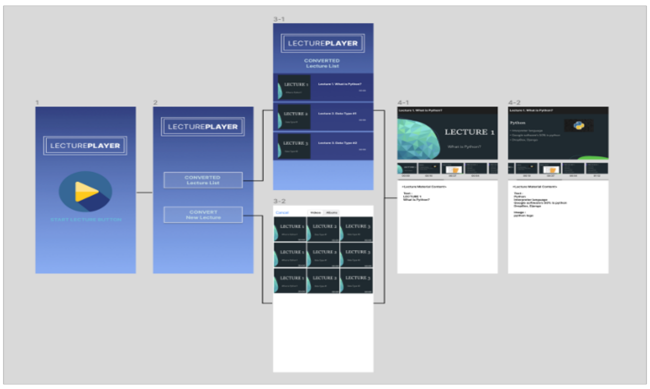

Fig. 3. The overall implementation screen of the automatic lecture video content commentator

3.2 Analysis and Commentary method of visual data in lecture videos by type

1) *Text*

After dividing the lecture video by scene, the existence of the text is first determined, and if it exists, the content is analyzed. To do this, we use the Google Cloud Vision API [23]. It provides a pre-trained machine learning model that assigns labels to images and quickly classifies them into millions of predefined categories. It supports several functions, such as text detection, logo detection, and label detection, among which it uses text detection. The text detection function performs the OCR function and structurally recognizes the text in the image in the order of page, block, and paragraph. The results are output in JSON form, which is parsed at the application server and stored in the JSON file containing information from the lecture video.

2) *Image (Picture)*

Images in lecture videos are one of the visual materials often used to help understand the contents of the lecture. However, they are a type of data that is difficult for visually impaired people to understand independently without external help. Inserting alternative text into the image can help that, but since many images do not have alternative text, we would like to create a description of the image using image captioning technology and provide it to the visually impaired. The Microsoft Azure cognitive computer vision API [24] uses AI models to recognize the image's main objects, scenes, and situations and automatically generate one sentence describing that. This API generates a description of the image and stores it in a JSON file structure containing information from the lecture video on the application server to provide it to visually impaired students.

3) *Diagram*

A diagram is an image showing the relationship between nodes or objects in the form of an arrow or connection line from one node to another, and in addition to text recognition, a process of analyzing the connection relationship between each node is required. In addition, protocols that explain the relationship between detected objects in a way that is easy for visually impaired people to understand also play an important role. Therefore, the following process is required to create a single sentence describing the diagram. First, we investigate the algorithmic approach that detects relationships between nodes in a diagram image. Second, we research how to effectively express the relationship between nodes in a diagram image in natural language to explain the diagram in an easy to understand.

Firstly, to analyze the relationship between nodes within a diagram image, we detect the text of the diagram image and the approximate type of diagram using the Google Cloud Vision API. Diagram images are characterized by the text being contained in bounding boxes. The text detection function of the API structurally analyzes the text in the image to recognize text in units of bounding boxes and outputs the box's location coordinates along with the text's information. Also, the label detection function of this API is used to detect the types of diagram images, such as horizontal, circular, and tree diagrams. The information of the detected type is included in the diagram description and provided to the visually impaired to help them understand the diagram. The arrow detection function is developed using Python by ourselves. The arrow is characterized by 4 or 5 sides of the convex shell and two more points not included in the convex shell. After properly preprocessing the image, find the shape that meets the arrow-shaped conditions by obtaining the number of convex shells and vertices of the shapes in the image. To determine the endpoint of an arrow, we use the method of finding the index of the arrow endpoints in an array of arrow vertices based on the two concave points of the arrow. After finishing the task, the relationship between the diagram nodes can be determined by comparing the text bounding box's position coordinates with the arrow endpoints' coordinates. This information produces a diagram description by connecting nodes into a single sentence in appropriate languages. Fig. 4 shows an example of the diagram analysis process.

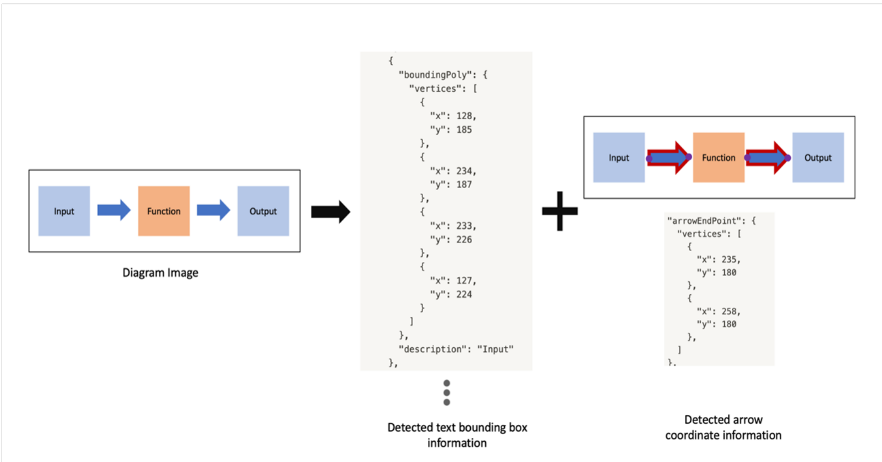

Fig. 4. Example of Diagram Analysis Process

4 Evaluation of usability of diagram commentary techniques

To convert a diagram into text, rules for reading the diagram are essential. This rule's setting is crucial to convey the exact meaning to the visually impaired student who cannot see the picture. In the United States, the diagram is described by referring to [17], a US version of the guidelines for explaining images of digital data. In this paper, we refer to this to create guidelines for explaining diagrams more effectively, divide the details of the explanation into three stages, examine which method helps visually impaired students the most, and apply it to developing the application that is more helpful to the visually impaired.

4.1 The method of the usability evaluation

1) *The Participants of the usability evaluation*

The usability evaluation participants consisted of three blind college students and six non-disabled college students. They consisted of students who had taken lectures at least once at the university. Visually impaired college students were recommended through the cooperation of the Korean Blind University Student Association. Although the number of visually impaired participants is small, this is generally inevitable in evaluating visually impaired people [25].

Before conducting the prototype test, visually impaired and non-disabled students who decided to participate in the usability evaluation were given a recording of the purpose of the study, privacy, and voluntary consent, and then agreed by voice. Also, we explained that recorded data are used only for research purposes and are discarded after all studies are completed.

2) *Usability evaluation tool*

An audio prototype was created for this usability evaluation. The application implementation still needs to be completed, and the implementation direction will vary by usability evaluation, so the audio prototype was produced first. The audio prototype was produced in the form of a voice file containing the description of the diagram by dividing it by the type of diagram and the degree of detail. There are four types of diagrams by difficulty: easy picture diagrams, circular diagrams, hierarchical/tree diagrams, and flow diagrams. A total of two examples were prepared with simple and complex examples per type, and in total, there are prototypes for eight diagrams. The degree of detail was divided into three stages: simple detailed, usually detailed, and very detailed. Therefore, three voice descriptions were prepared for each of the eight diagrams. The diagram description rule of the third step is shown in Table 1, and as an example, the description of Fig. 5 is shown in Table 2.

Table 1. Diagram Description generation Rules

The degree of detail	Description rules
Simple detailed	Only the text in the diagram is read in order, and the word "next" is inserted between the texts.
Usually detailed	Present the title of the diagram at the beginning of the description. It wraps the text in the diagram with natural sentences and connects the sentences with the words "Follow the arrow."
Very detailed	Describe the title of the diagram image and the shape of the diagram at the beginning of the description of the diagram. Wrap the text in the diagram in a natural sentence and connect the sentences by specifying the letters at the beginning and the end of the arrow.

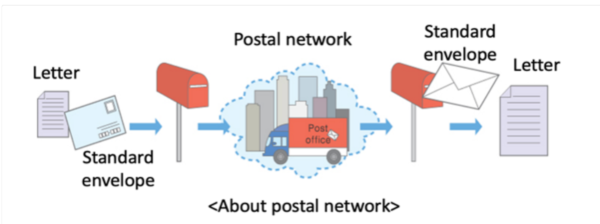

Fig. 5. Diagram for description examples

Table 2. Example Diagram description for Fig.5

The degree of detail	Description rules
Simple detailed	About postal network. Letter. Standard envelope. Next. Postal network. Next. Standard envelope. Next. Letter. End.
Usually detailed	The title is “About postal network”. There are “Letter,” “Standard envelope” letter and two paper drawings. Follow the arrow, there is mailbox drawing. Follow the arrow, there are “Postal network” letter and a mail carrier driving through the city drawing. Follow the arrow, there are “Standard envelope” letter and mailbox and paper drawing. Follow the arrow, there are “Letter” letter and paper drawing. The end.
Very detailed	The title of the diagram is “About postal network.” Diagrams are easy picture diagrams that progress horizontally with arrows. At the beginning of the diagram, there are "Letter", "Standard Envelope" letters and two paper drawings. Follow the arrow in "Letter", "Standard Envelope", there is mailbox drawing. Follow the arrow, there are “Postal network” letter and a mail carrier driving through the city drawing. Follow the arrow in “Postal network,” there are “Standard envelope” letter and mailbox and paper drawing. Follow the arrow in “Standard envelope,” there are “Letter” letter and paper drawing. The end.

The examples of diagrams used in the experiment were collected from lecture materials uploaded to the online education platform of the university of this research team. Eight diagrams suitable for prototyping were selected among lectures agreed to be used for the study. [17] was referred to when producing the diagram description.

The Likert 5-point scale [26] was used to determine the understanding and satisfaction of the description of one diagram. In addition, to find out how much the audio description affects the understanding of the diagram, three quizzes were prepared at the end of the diagram, and the researcher counted the number of correct answers. The quiz was prepared as a short answer question that can be answered when the explanation is accurately understood regardless of prior knowledge of the diagram.

3) *The procedure of Usability evaluation*

The prototype test interview was conducted for approximately five days, from December 22, 2022, to December 27, 2022. To prevent the spread of the virus due to the epidemic of COVID-19, it was conducted through the video conference program "Zoom." Each interview took about an hour for each participant, and the prototype was played through the sound-sharing function of the video conference program. The interview was conducted by following orders reading the consent form to participate in the study and guiding precautions, simulation tests before starting, understanding and quiz questions after playing the diagram description recording, final questions about all diagram descriptions, and finishing the test interview.

For each participant, a description of one detail stage was randomly given to prevent situations in which a particular stage is always explained later and has a higher understanding. In addition, the type of diagram that starts for each participant was randomized so that a particular type would always be the first to start the test or the last to do so that it would not affect understanding.

Before playing the diagram description voice file, the degree of the detail of the file informed the participants. After tests are completed, it can help the study participants remember the degree of detail to answer the question that asks the most satisfactory of the three degrees of detail. After the voice file was finished, they were asked about understanding, three quizzes, and satisfaction. A spare quiz was prepared and conducted to control the quiz's difficulty according to the test participants' responses. After the tests on eight diagrams were completed, a final question was asked about their feelings about the prototype test.

4.2 The Result of Usability Evaluation

Fig. 6 shows the results of averaging the participants' understanding, quiz, and satisfaction scores for each detailed stage of the diagram description as a result of the prototype test. In the test, understanding and satisfaction were out of 5, but the quiz score was out of 3, so we converted it into out of 5 and drew a graph.

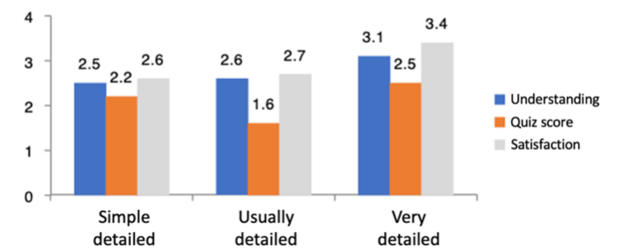

Fig. 6. Average graph of scores according to the degree of detail of the diagram commentary

As seen from Fig. 5, the commentary in explaining the diagram in the very detail scored higher in terms of understanding and content quiz scores than other commentaries. There is no significant difference in scores between a simple description of detail and an explanation of a usual step. The automatic diagram commentator for the visually impaired aims to help students understand various diagrams most efficiently. If only one type of diagram commentary is provided, it may be difficult for students to understand a diagram in which other types of commentary are more effective. Therefore, more diagrams can be effectively understood if the user can select and listen to the explanation depending on the type of diagram between a simple detailed explanation and a very detailed explanation.

5 Conclusion

In this paper, a diagram explanation function design among automatic lecture video content commentators was presented to help visually impaired students better understand online lectures. It focused on explaining the visual materials of the lecture video in the text as much as possible and delivering them to users, thereby improving the learning efficiency for the visually impaired. Descriptions of text, images, and diagrams among visual materials were automatically generated and provided, and usability evaluations were conducted to find ways to help students with visual impairments the most for system development. We are confident that this can provide an opportunity to create an independent learning environment for visually impaired students. In addition, adding a system that automatically interprets more diverse visual materials, such as tables and mathematical graphs, will bring the application closer to its final goal.

Acknowledgements This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2023-RS-2022-00156299) supervised by the IITP(Institute of Information \& Communications Technology Planning \& Evaluation). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.2022R1F1A1063408). This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation)"(2022-0-01087). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022H1D8A3037394).

References

1. M. Wende, S. Bulut, T. Giese and R. Anderl, "Online Lectures – Strategies for Applying Advanced Digital Media for Future Higher Education," 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Takamatsu, Japan, 2020, pp. 311-318, doi: 10.1109/TALE48869.2020.9368353.
2. U. Akcil, "The use of mobile learning for visually impaired learners school in tolerance education contents," *Quality and Quantity*, vol. 52, pp. 969-982, December 2018.
3. Y. S. Shin and H. M. Cha, "A Study on the Teaching Experience of Visually Impaired School Teachers According to the Corona 19 Situation," *The Korean Journal of Visual Impairment*, vol. 36, no. 4, pp. 147-168, 2020.
4. M. del Pilar Oviedo-Cáceres, K. N. Arias-Pineda, M. del Rosario Yepes-Camacho, *et al.*, "Covid-19 Pandemic: Experiences Of People With Visual Impairment", *Investigacion y Educacion en Enfermeria*, vol. 39, no. 1, 2021.
5. D. Y. Park and S. B. Lim, "Video Screen Commentary System Supporting Online Learning of Visually Impaired Students," *Lecture Notes in Computer Science*, vol. 13342, pp. 360 – 368, 2022.
6. D. Y. Park, "Voice Commentary System to Increase Understanding of Video Lectures for Visually Impaired Students," *Master's Thesis*, Sookmyung women's university, Seoul, Republic of Korea, 2023.
7. J. H. Park, J. W. Lee, J. S. Um, *et al.*, "Korean language math-to-speech rules for digital books for people with reading disabilities and their usability evaluation," *Journal of Supercomputing*, vol. 77, no. 6, pp. 6381-6407, 2021.
8. H. W. Na, J. Yook, S. Y. Dong. "Automatic Generation of Mathematical Graph Descriptions for Students with Visual Impairments," *Journal of the Korean Society of Rehabilitation and Welfare Engineering*, 79-79, 2020.
9. R. Shrestha, "A Neural Network Model and Framework for an Automatic Evaluation of Image Descriptions based on NCAM Image Accessibility Guidelines," *ACM International Conference Proceeding Series*, pp. 68, 2021.

10. J. L. Davenport, D Yaron, D. Klahr, *et al.*, "When do diagrams enhance learning? A framework for designing relevant representations", *Computer-Supported Collaborative Learning Conference*, CSCL, pp. 191, 2008.
11. Survey of Web Accessibility Practitioners Results. https://webaim.org/projects/practitioner-survey/ (2014), last accessed 2020/2/22
12. Korean institute for health and social affairs, "Survey on the Status of Persons with Disabilities in 2020," http://www.kbuwel.or.kr/Board/FileRoom/Detail?page=1&contentSeq=1212321, last accessed 2023/5/30
13. J. Ismaili and E. H. O. Ibrahimi, "Mobile learning as alternative to assistive technology devices for special needs students," *Education and Information Technologies*, vol. 22, no. 3. Springer Science and Business Media LLC, pp. 883–899, 19-Jan-2016.
14. M. Al-Razgan, S. Almoaiqel, N. Alrajhi, *et al.,* "A systematic literature review on the usability of mobile applications for visually impaired users", *PeerJ Computer Science*, vol. 7, 2021.
15. M. Shoaib, S. Khan, D. Fitzpatrick, *et al.*, "A mobile e-learning application for enhancement of basic mathematical skills in visually impaired children," *Universal Access in the Information Society*, 2023.
16. The DIAGRAM Center, "Image Description Guidelines," June 2015.
17. Flutter, https://flutter.dev/., last accessed: 2023/7/6.
18. Flask, https://flask.palletsprojects.com/en/2.3.x/ , last accessed: 2023/7/6.
19. Python, https://www.python.org/ , last accessed: 2023/7/6.
20. SQLite, https://sqlite.org/index.html , last accessed: 2023/7/6.
21. PySceneDetect API, https://pyscenedetect.readthedocs.io/en/latest/reference/python-api/ , last accessed: 2023/5/30.
22. Google Cloud Vision API, https://cloud.google.com/vision?hl=ko , last accessed: 2023/7/6.
23. Microsoft Azure cognitive computer vision API, https://azure.microsoft.com/ko-kr/ , last accessed: 2023/7/6.
24. J. Lazar, J. H. Feng, H. Hochheiser, Research methods in human- computer interaction. Wiley, London, 2010.
25. R. Likert, "A technique for the measurement of attitudes." *Archives of psychology*, 1932.