Development of an electronic book accessibility standard for physically challenged individuals and deduction of a production guideline

Joo Hyun Park, Hyun-Young Kim, Soon-Bum Lim*

Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul, Korea

ARTICLE INFO

ABSTRACT

Unlike popular web-based internet and mobile applications, electronic books and similar products have lower distribution and less accessibility, making them difficult for vulnerable social groups to utilize. Various institutions have proposed electronic publication (EPUB) accessibility guidelines, based on the IDEF modeling language, to provide accessible content for electronic publications. However, these are merely guidelines for securing accessibility of common electronic books. They are not geared toward the visually impaired persons. Moreover, even if accessibility standards exist, it becomes even more difficult to then find examination tools that can fully verify standards compliance. This study establishes an electronic book accessibility standard for the physically challenged individuals, based on the EPUB 3.0 accessibility guideline. We developed an automatic/semi-automatic examination tool that can test the standard. We linked the SIGIL electronic book reader, having the highest market share in Korea, to an examination tool to verify error detection performance. Lastly, an accessibility test was performed on 50 electronic books, commercially used in the Korean electronic book market, to assess which accessibility standard domain is the most problematic. Thus, we propose a guideline on matters that should be considered when writing electronic books in Korea.

1. Introduction

Korea's information and communication technology development speed may be the highest in the world. However, this is based only on external growth. Inside Korea, there ought to be an environment where the 2.5 million physically challenged individuals, who make of 5% of the entire population, can enjoy their fair share of electronically published information [1]. Unlike widely popular web-based internet and mobile applications, electronic books and similar products have lower distribution and less accessibility, making them difficult for vulnerable social groups to utilize. In our current era of informatization, any gap in information creates wider gaps in other domains, including education and income [2].

Physically challenged individuals are defined as not only persons with physical disabilities but also situations in which reading activity is interrupted by external environmental factors. It includes, for example, all blind and low vision due to inherited, acquired factors, and who are not free to use the visual, such as when driving, cooking, and so on. Also included, reading disabled people including dyslexia and illiteracy.

A few electronic publication (EPUB) accessibility guidelines exist today, including the International Digital Publication Forum (IDPF), the Book Industry Study Group (BISG), and the Digital Accessible

Information SYstem (DAISY) consortium. These guidelines provide required accessibility checkpoints and rules that must be adhered to by developers that create electronic books. From the perspective of publishers or writers who are not accustomed to web technologies (e.g., HTML, CSS, and EPUB), substantial time and money is wasted creating adhering to these guidelines, because they believe doing so requires foreign-language fluency and significant manual labor. This opinion inhibits the mass production of accessible EPUBs. According to [3], in Korea, a publisher first converts the EPUB from a PDF or ADOBE InDesign file, inserts some multimedia files, adds interacting JavaScript, and then distributes the PDF without essential options required for the user. This demonstrates that they have no interest in creating a truly accessible EPUB file. Upon examining the EPUB files that has been donated to the National Library of Korea, 95% of needy users found that these files do not have the functions they require.

Furthermore, extant EPUB accessibility guidelines include use of electronic books by the physically challenged individuals, but they do not specify accessibility for them. Structural accessibility standards for electronic books must be designed such that all the physically challenged individuals enjoy accessibility [4]. However, they must also be meaningfully accessible. Not only is it difficult to find electronic book accessibility standards for the physically challenged individuals, but

even when such standards exist, it is difficult to find an examination tool that can fully verify them.

The HTML Validator, an automatic accessibility tester for web content, tests tags in hypertext documents. However, it only tests the structural aspects of a markup language, meaning it cannot test all potential barriers. This problem can also be found in the EPUB field. EpubCheck, which tests the accessibility of EPUB documents, reviews the structural issues of an EPUB, but it cannot review the semantic issues that occur. For example, when determining whether the page numbers match the pages of the actual book, it requires human intervention to decide.

In this study, an electronic book accessibility standard of 154 accessibility items, specialized for individuals with reading disabilities, is established, based on the EPUB accessibility standards of various organizations. Additionally, a verification tool is developed to test this standard. The verification tool consists of a two-stage verification system. In the first stage, the automatic verification system disassembles and analyzes the EPUB structure and performs automatic verification based on 48 component accessibility items. In the second stage, the semi-automatic verification system checks the accessibility of each component via a human specialist. Lastly, the verification tool is used to evaluate the accessibility of electronic books available in the Korean market for the physically challenged individuals. The most problematic area of accessibility is assessed to propose a guideline that should be considered when writing an electronic book. Moreover, the electronic book accessibility specifications the physically challenged individuals established through this study were established as a Telecommunication Technology Association (TTA) standard document in Korea.

2. Related studies

Accessibility standards include those from the World Wide Web Consortium (W3C), the IDPF, the DAISY Consortium, and the Digital Image and Graphic Resources for Accessible Materials (DIAGRAM) Center. The IDPF published the "EPUB Accessibility Guidelines [5]" and their associated standardization tasks [6]. The W3C published the "Web Content Accessibility Guidelines (WCAG) 2.0 [7]" and the "Accessible Rich Internet Applications (WAI-ARIA) 1.0 [8]," containing web accessibility specifications and guidelines that accommodate HTML, CSS, and JavaScript. The W3C and IDPF have provided separate specifications and definitions tailored for electronic books, including structure and navigation, text-to-speech, and designation of a logical reading order.

The DAISY Consortium has provided accessibility factors, defined in the Digital Talking Book specification, addressing accessibility [9]. These specifications are divided into content accessibility rules, authoring tool accessibility rules, user agent (i.e., user application) accessibility rules, and content accessibility guidelines, which are nearly identical to those of the IDPF.

Web accessibility testers, such as those built by Greg Gay, are open and conversational and provide web accessibility and user definitions [10]. When the web accessibility checker tool is used, the reviewer can interact with the system while they review accessibility, allowing decisions to be made on issues that cannot be fully identified by a system. This is very similar to the semi-automatic verification process that address problematic points at web viewer. If accessibility rules are not satisfied, the web editor send a notice to make the editor changed those points so that this point can also be changed from the web editor. This tester only works for web content and not for EPUB content.

There are no libraries on the Korean market that related to electronic book accessibility, including e-books composed of PDFs or images format. However, libraries that can verify that EPUB specifications are satisfied have been presented by IDPF (e.g., EpubCheck) [11]. This library support both EPUB 2.0 and 3.0 specifications, and it automatically verify the structure of each version: open container format, an EPUB specifications factor; open packaging format (OPF); open publication structure; and EPUB publication 3.0 markup. It also verifies the consistency of the internal reference model.

3. Development of an electronic book accessibility standard for the physically challenged individuals

An electronic book accessibility standard for the physically challenged individuals should include structural accessibility standards and standards for accessible electronic-book content for readers. These accessibility standards have leverage existing EPUB accessibility guidelines on various organizations. Having reviewed the " EPUB 3.0 Accessibility Guideline " from IDPF, the " Quick Start Guide to Accessible Publishing " from BISG [12], " Top Tips for Generating Accessible EPUB 3.0 Files " from DIAGRAM Center [13], WCAG 2.0, and WAI-ARIA 1.0 guidelines from W3C, we extracted 154 accessibilities authoring guidelines, including term definitions, separate style files, bold tags, and tables of content .

The accessibility verification items extracted from each guideline are as follows. The "EPUB 3.0 Accessibility Guideline" was designed to help create accessible EPUB 3.0 content. It includes rules for how to fulfill requirements and a quality assurance checklist. It also includes 63 categories from the semantic expression of a logical reading order and interactions with the canvas and script. We extracted parts in this guideline that can be implemented with an automatic tester.

The "BISG Quick Start Guide to Accessible Publishing" is a model case that creates accessible digital content and adheres to international standards. Further, it explains why and how to write, distribute, and indicate accessible digital content and provides an overview on the importance of accessibility. This guideline is very similar to the EPUB 3.0 accessibility guideline and is connected to the "Image Guideline for EPUB 3.0″ from DIAGRAM. We involve checking for text that replaces images or image explanations, and accessible metadata.

WCAG 2.0 addresses recommendations of a wide scope to increase the accessibility of web content for users that are visually challenged, suffer hearing difficulties, learning disabilities, cognitive disabilities, limited movements, or linguistic difficulties, or a combination of these. EPUB provides the basics of web content, such as HTML and CSS. Therefore, each factor and each page of an EPUB electronic publication must adhere to WCAG specifications. We had the checker utilize a WCAG at Level A success standards.

To help facilitate the appropriate information being delivered to the physically challenged individuals, semantic information, regarding widgets, structure, and behaviors, requires web content accessibility. WAI-ARIA 1.0 defines accessible user interface factors and provides an ontology regarding the roles, status, and attributes that can be used to improve accessibility and interaction of web content and application programs. Such requirements can trigger interaction with readers of electronic books. We matched these functions to a semi-automatic test process connected to a web viewer and editor.

The accessibility standard of the electronic book for the physically challenged individuals which we derived has established as the Korean standard through the resolution of the Korea Information and Communications Technology Association (TTA)'s IT standardization meeting. TTA is an organization that establishes standardization in Korea in connection with standardization trends such as ITU-T. Part 1 contains authoring rules, and Part 2 contains certification standards, which provide Accessibility Guidelines for Reading Disabled People. Fig. 1 shows the process of developing an electronic book accessibility standard for the physically challenged individuals.

Fig. 1. The process of developing an electronic book accessibility standard for the physically challenged individuals.

4. Accessibility verification system design and implementation results

4.1. Overview of the accessibility verification systems

The accessibility verification tool consists of a two-stage verification system. In the first stage, the EPUB structure is decomposed and accessibility is verified from the PC version. In the second stage, the overall accessibility of each component is verified by a human specialist for the web version. The first stage is called "the automatic test" whereas the second stage is called "the semi-automatic test" The automatic part of the first stage is linked to SIGIL so that verification can be performed on the SIGIL screen once an electric book is written.

4.2. System configuration and process

The system runs the automatic verification system during the first stage and disassembles the EPUB structure automatically and checks all 48 accessibility test points for each factor, cross-referencing relationships each time an EPUB file is obtained. Errors from verification results are shown in the form of a table with the location, row, and column number of each problem in the component file of the EPUB, as shown in Fig. 2.

Furthermore, the correct specifications are shown so that the editor can correct the problematic point. The 48 accessibility standards exclude items that are unrelated to content accessibility, such as the scope of the viewer accessibility, EPUB semantic term regulations, ARIA role attributes, and EPUB semantic structure definitions.

Once the first stage is complete, the EPUB file is uploaded to the web system, and the semi-automatic test that is the second stage begins. This involves processing the remaining 106 checkpoints, verifying the overall accessibility of each component. Each time the system obtains an EPUB file, it automatically selects a specialist who is registered in the pool of accessibility specialists and sends a verification request via e-mail. The specialist logs into the system and begin the second stage process. The semi-automatic checker searches for error points that may cause accessibility issues and displays them on the web viewer. The specialist manually checks these points for issues. The original copy of the final defect report collected from the specialist is sent to the publisher. Lastly, the publisher and editor have the opportunity to correct errors. Then, the process may begin all over again. Fig. 3 shows the overall system process.

4.3. Implementation of verification system and results

To implement the automatic verification library, the EPUB structure must first be disassembled. Thus, a library that processes the EPUB ZIP structure is required. Additionally, an understanding of XML/XHTML,

CSS, SMIL, and SVG syntax is required, and their accessibility factors must be verified accordingly. Therefore, in this paper, an open library is used to implement the relevant fundamental tasks. To link the automatic verification library to the actual electronic book authoring tool, or other tools related to electronic books, it must be linked to a verified product. Therefore, we processed it in the JSON format so that delivery of the verified product, the syntax and location of the error, explanation for error correction, etc., could occur smoothly without requiring additional steps.

For an understanding of the EPUB structure, the implemented method of the EpubCheck library was referenced and used by parsing the OPF structure definition files. Furthermore, actual EPUB components were disassembled to assess the relationship between components and the location of each component file included in the structure definition.

As shown in Fig. 4, the PC version was developed using C# and only runs on a Windows OS. It has six menus: Open (EPUB), Report (Unresolved Report), Batch Processing (processing multiple EPUB files at a designated time), Environmental Settings, Help, and Exit. It also has three modules: EPUB disassembly module, accessibility test module, and the EPUB editor.

The web version, shown in Fig. 5, was developed using Java and Maven. It consists of three parts: Manager (all specialists and procedural management, statistical data review), Publisher (checks status of EPUB files), and Specialist (takes part in tasks during the second stage). It also includes five modules: EPUB management module, accessibility test module, EPUB viewer, process management module, and personal management module.

To check the completeness and reliability of the functions of the completed accessibility verification library, a total of five sample EPUB files were created, including all a file 48 errors and random number of errors. The sample EPUB files presented their respective error explanation syntax from the IDPF EPUB Accessibility Guideline. Then, files were created to include all correctly and incorrectly example.

5. Experiment results and considerations

This section assesses the EPUB accessibility issues in Korea by verifying the completeness of the proposed electronic book accessibility verification tools for the physically challenged individuals and by evaluating electronic books used in the verification system.

5.1. Completeness of the automatic verification system

The automatic accessibility verification system was connected to SIGIL, the most commonly used EPUB authoring tool in Korea, to inspect accessibility verification functions.

File	Row	Column	Message
OPS/c01.xml	9	38	Already has background-image attribute
OPS/c01.xml	12	31	No \<aside\> Tag
OPS/c01.xml	20	42	No \<lang\>
OPS/c01.xml	23	31	No \<xml:lang\>

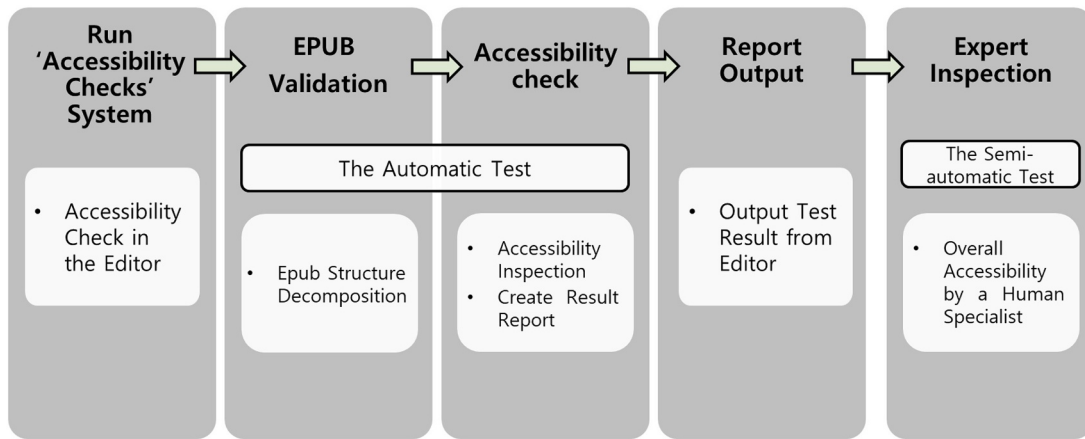**Fig. 2.** Screenshot of Errors of Verification Result.

Fig. 3. Overall System Process.

The completeness of the automatic verification system was verified through efficiency comparison. Efficiency comparison was measured by success rate and time. We provided the same Epub sample files to the participating systems and participants and asked them to find points that violated the accessibility verification criteria. Five participants participated in the test, and all of them have experience in making EPUB files. 5 sample EPUB files are used for verification, each of files contains 48, 30, 23, 15, and 10 error items. The participants were instructed to manually inspect five sample files without informing them of the number of errors in each sample file.

In the results, the system took an average of 1.14 min, 0.83 min, 0.71 min, 0.53 min and 0.53 min to validate each sample file, while the participants took an average of 68.7 min, 66.5 min, 64.4 min, 63.7 min and 60.8 min. In the success rate results, the system detected 100% errors in all files, while the participants detected an average 68.7%, 65.5%, 64.4%, 63.7% and 60.8% for each file. Based on these results, we can find that system efficiency is higher than manual method (Fig. 6).

Therefore, the Automatic Verification System confirmed that the inspection time was significantly reduced compared to the manual
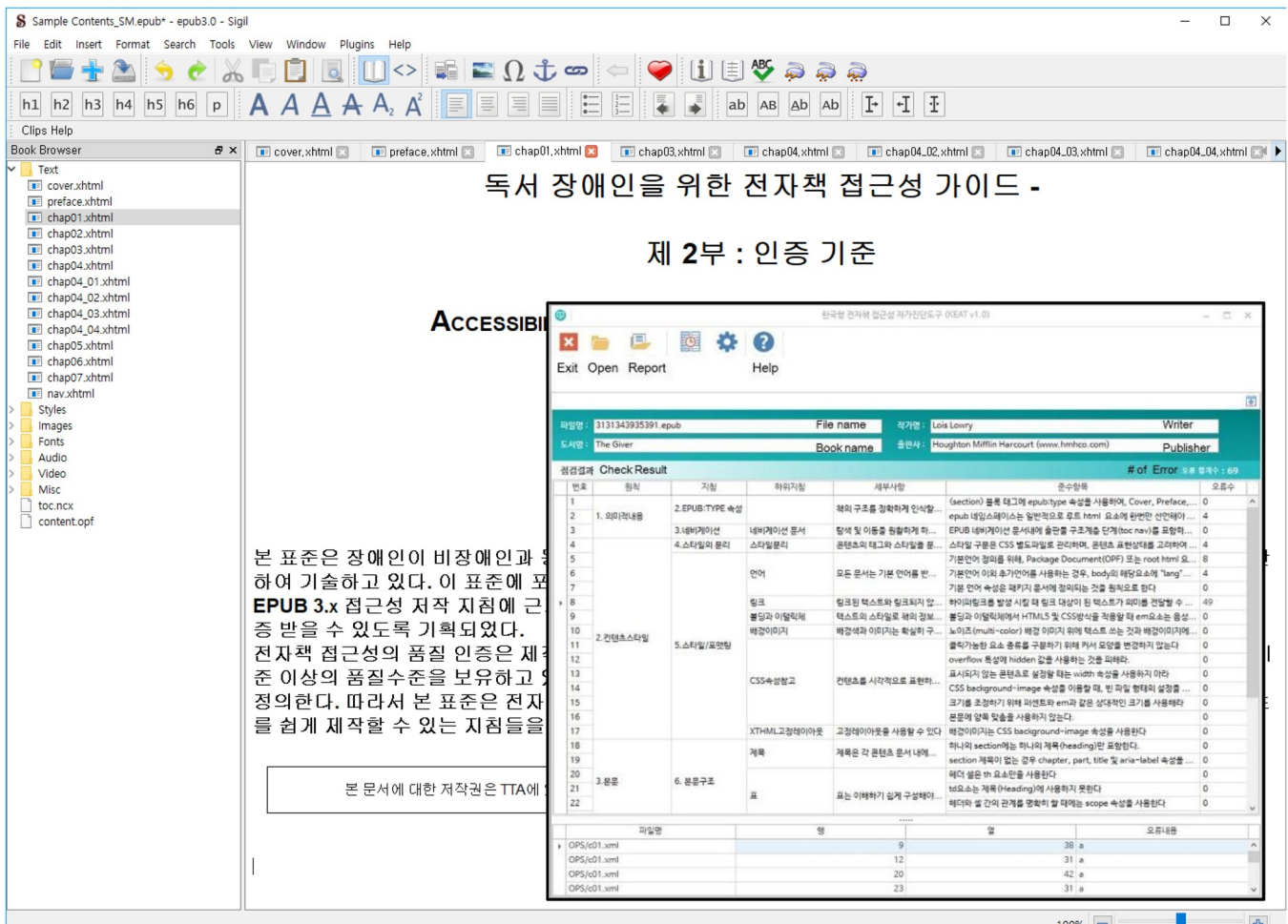

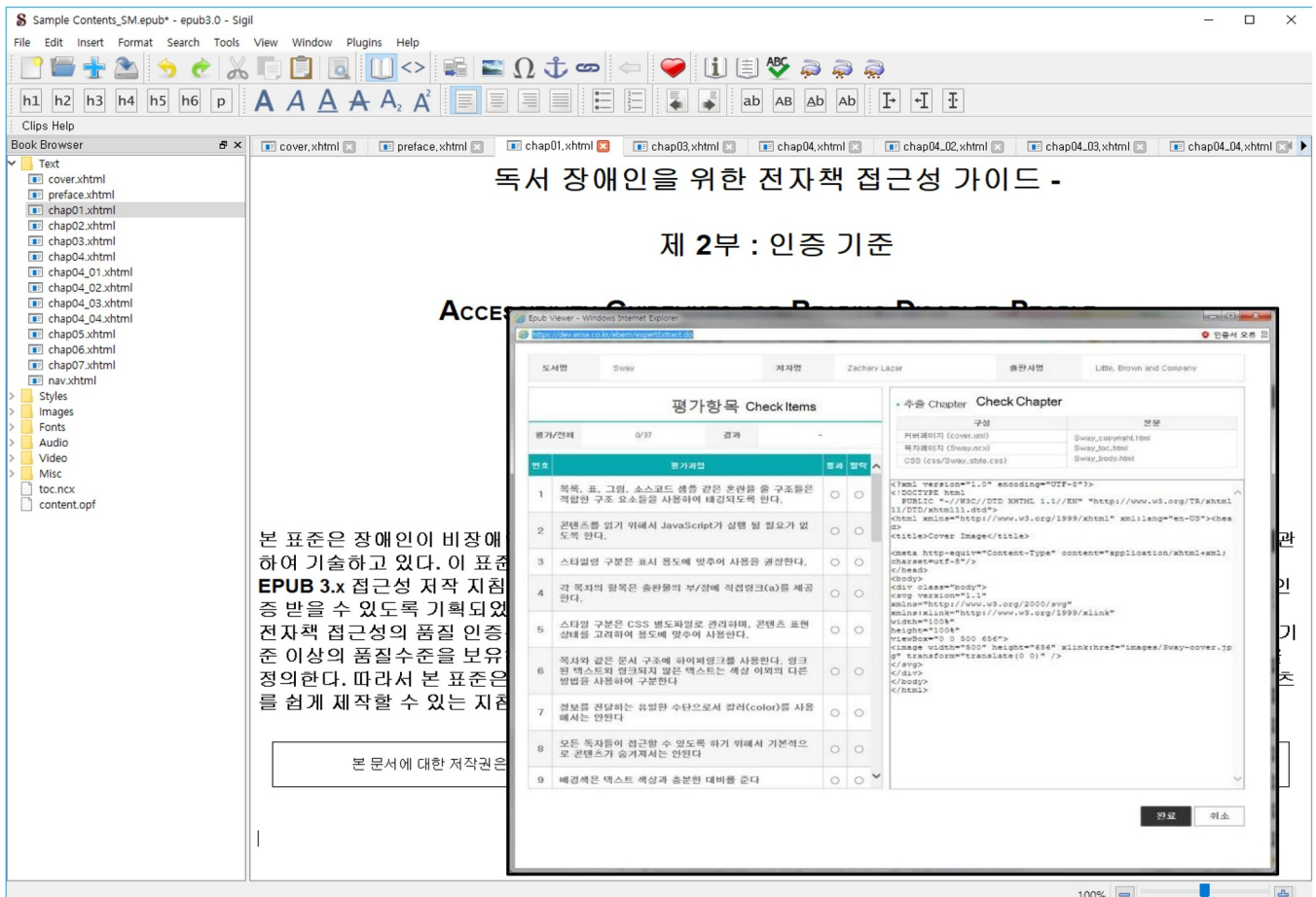**Fig. 4.** Screenshot of PC Version of Dual-Layer System.

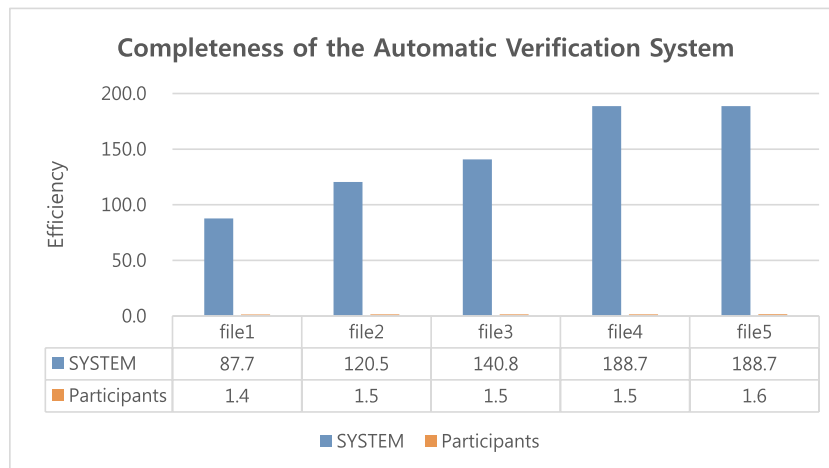**Fig. 5.** Screenshot of Web Version of Dual-Layer System.

Fig. 6. Efficiency comparison bar graph.

inspection, and that the point of violation of accessibility standards was accurately found.

5.2. Deducing guidelines through a field assessment

To deduce a guideline for creating electronic books that comply with the accessibility standards for the physically challenged individuals, 50 Korean EPUB files, which were donated to the National Library of Korea, were selected. Those files were surveyed on how closely they adhered to accessibility specifications. Publishers in Korea typically used "SIGIL (Hendricks)," which is a free, open source editing software for electronic books in EPUB format. It was presented after the EPUB format was on the market. This software supports some accessibility features of the current version. However, there were an average of 148 accessibility defects in each of the 50 sample EPUB files. We believe that these sample files were generated in a previous version of SIGIL that does not support accessibility. Moreover, because the electronic book market in Korea does not yet cover EPUB 3.0 multimedia, tables, SVG, TTS, and various other items, most of the errors occurred in the text and style processing areas, including definition of language in

Table 1
Accessibility errors from 50 commercial EPUB files.

	Detected Accessibility Standards	# of errors	Ratio
1	The ⟨figurecaption⟩ element is located at the beginning or end of the figure element.	6	0.08
2	Use relative sizes units such as ems and percentages to facilitate scaling.	28	0.38
3	Avoid aligning the text, as uneven spacing between words can reduce readability.	76	1.03
4	CSS properties for italics and bolding should be used each time the usage is presentational.	180	2.43
5	Separate markup and style by keeping CSS as a separate file.	435	5.88
6	Each section cannot have more than one heading.	471	6.37
7	The hyperlinked text should have meaning in the text itself, not the color. If the linked text cannot convey meaning, you must add the title attribute to the tag that describes the attribute.	459	6.20
8	The language attribute should be used to identify the language for the content.	1499	20.26
9	When using the epub: type attribute in a content document, must declare the epub namespace in that element or its parent element.	921	12.44
10	Meaningful images in publications should always include alt text via the "alt" attribute.	432	5.84
11	To define the default language in an XHTML document, must add the "lang" and "xml: lang" attributes to the ⟨html⟩ element. The language added here must match one of the languages specified in the package document metadata dc: language	2893	39.09
	Total	7400	100.00

Korean and English mixed phrases, heading tags designations, such as h1/h2 for a page or section, end alignments, styling file separations, etc. Thus, we can confirm that the function of the automatic verifier works correctly through this experiment. Table 1 shows the quantity of each error found through the automatic accessibility verification on 50 sample files and the proportion from the entire error count. Also, the "Detected Accessibility Standards" column in the table lists 154 accessibility guidelines.

Accessibility issues focus on the following seven areas for EPUB files in the Korean market.

- When the basic language of a XHTML document is designated, the "lang" and "xml:lang" property must be attached to the root html element. This accounts for about 39% of all errors.
- For publications with multiple languages, it is best to designate the language of each document at all times, so that the correct rendering can be made. This accounts for about 20% of all errors.
- If the epub: type attribute is used in the document, the epub namespace must be declared in either the element or its parent including the property. This accounts for about 12% of all errors.
- Each section cannot have more than one heading. Each heading element must represented by a single heading. This accounts for about 6% of all errors.
- When creating a hyperlink, the text inside the link should provide the entire context of the linked context or the link should have alternative text. This accounts for about 6% of all errors.
- Separating the style from the markup should result in not only maintaining the CSS in the file separate from the markup, but also recognizing that the markup should convey meaning useful to all readers.
- Key images for understanding the publication must always include alternative text in the alt property. This accounts for about 6% of all errors.

The above seven key items refer to the central issues that must be considered when writing for accessibility in Korea epub market. Fortunately, these issues can be revised if they are detected in our accessibility verification system. Because our system can detect errors automatically, this ultimately implies that authors will experience a reduction in production time and, because manual tasks will no longer be required.

6. Conclusion

This study established an electronic accessibility standard for the physically challenged individuals based on the EPUB 3.0 accessibilities rules from the IDPF and other institutions, and developed an automatic/semi-automatic verification tool that can test the standard. The verification tool involved a structural analysis of EPUB files with an automatic verification based on 48 accessibility items on the components.

Moreover, it detected accessibility errors and provided precise specifications for the errors (i.e., row and column number from the code level) and a guide for revising errors (i.e., example of commands to be revised/replaced) to guide the editor to correct the automatically detected errors. Facilitating accessibility verifications in electronic books, which grow with the paper book market, is expected to help reduce time and costs required to produce accessible electronic books. The functional assessment of the verification tool was performed with SIGIL to confirm the system's error detection performance. An accessibility test was conducted on 50 electronic books in the Korean electronic book market to assess which accessibility factors are the most problematic. These issues were presented as a guideline for consideration when authoring electronic books in Korea. The accessibility standard for the physically challenged individuals, which is the product of this study, has been established as the standard document for the TTA of Korea.

The completeness of the proposed electronic book accessibility standard for the physically challenged individuals may be evaluated based on whether or not the established accessibility standards are properly used in industry. This standard is currently established as two parts of the TTA standards (i.e., TTAK.KO-10.0905-Part1 and TTAK.KO-10.0905-Part2) and is supplying on site. Because it is still in the early stages of distribution, evaluating completeness at this point is premature. Hence, this evaluation should be updated with further assessments and discussions once the accessibility standards have spread through the entire electronic book industry in the future.

Moreover, to improve understanding for individuals with reading disabilities in terms of electronic books, it is necessary to support and use more multimodal methods of more diverse multimedia than the electronic books currently on the market. However, as the multimedia and multimodal ratio increases, more complex accessibility functions should be implemented. Thus, the electronic book accessibility verification tool will require continuous upgrades. The manual verification will only exacerbate the problem, more research must be conducted on natural language processing, similar EPUB syntax analysis, and semantic analysis-based machine learning techniques to improve the automated performance of automatic verification tools.

Acknowledgments

References

[1] S.S. Noh, Trends and countermeasures of web accessibility international standard revision, Proceeding of Information Accessibility Technology Conference, 2016, pp. 13–35.
[2] J.H. Park, K.H. Lee, J.W. Lee, S.B. Lim, Design of voice annotation system to enhance the quality of life for visually-impaired readers, Asian Life Sci. 11 (2015)

109–121.

[3] H.Y. Kim, J.H. Park, S.B. Lim, Automatic and semi-automatic 2-tier Check System for EPUB accessibility, J. Technol. Persons Disabil. 5 (2017) 29–39.

[4] K.H. Lee, T.E. Kim, J.W. Lee, S.B. Lim, A design of mobile e-book Viewer interface for the reading disabled people, J. Korea Multimed. Soc. 16 (2013) 100–107.

[5] IDPF, EPUB 3 Accessibility Guidelines. https://idpf.github.io/a11y-guidelines/ (Accessed 30 December 2017).

[6] IDPF, EPUB Accessibility 1.0. http://www.idpf.org/epub/a11y/accessibility.html (Accessed 30 December 2017).

[7] World Wide Web Consortium, "Web Content Accessibility Guidelines (WCAG) 2.0, https://www.w3.org/TR/WCAG20/, (2008) , Accessed date: 23 January 2018.

[8] World Wide Web Consortium, Accessible Rich Internet Applications (WAI-ARIA) 1.0, https://www.w3.org/TR/wai-aria/, (2014) , Accessed date: 24 April 2018.

[9] ANSI/NISO Z39, 86-2005 (R2012) Revision of ANSI/NISO Z39.86-2002,

Specifications for the Digital Talking Book, http://www.daisy.org/z3986/2005/Z3986-2005.html?q=z3986/2005/z3986-2005.html, (2005) , Accessed date: 24 April 2018.

[10] Greg Gay, Cindy Qi Li, A Checker, Open, interactive, customizable, web accessibility checking, Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility, 201023.

[11] EPUB 3 Community Group, EpubCheck https://github.com/IDPF/epubcheck, (2012) , Accessed date: 2 January 2018.

[12] Book Industry Study Group, BISG quick start guide to accessible publishing, https://inclusivepublishing.org/publisher/bisg-quick-start-guide/ (Accessed 24 April 2018).

[13] DIAGRAM CENTER, Top Tips for Creating Accessible EPUB3 Files, http://diagramcenter.org/54-9-tips-for-creating-accessible-epub-3-files.html (Accessed 24 April 2018).