

An automatized method based on LaTeX for the realization of accessible PDF documents containing formulae

Original

An automatized method based on LaTeX for the realization of accessible PDF documents containing formulae / Armano, Tiziana; Capietto, Anna; Coriasco, Sandro; Murru, Nadir; Ruighi, Alice; Taranto, Eugenia. - 10896:(2018), pp. 583-589. (Intervento presentato al convegno International Conference on Computers Helping People with Special Needs).

Availability:

This version is available at: 11583/2719239 since: 2018-12-03T12:12:14Z

Publisher:

Springer

Published

DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

Springer postprint/Author's Accepted Manuscript

(Article begins on next page)

An automatized method based on LaTeX for the realization of accessible PDF documents containing formulae

Tiziana Armano¹, Anna Capietto¹, Sandro Coriasco¹, Nadir Murru¹, Alice Ruighi¹, and Eugenia Taranto¹

¹Department of Mathematics, University of Turin,
Via Carlo Alberto 10, 10121 Torino, Italy

{tiziana.armano,anna.capietto,sandro.coriasco,nadir.murru,alice.ruighi,
eugenia.taranto}@unito.it

Abstract. Assistive technologies for visually impaired people (screen readers and braille displays) perform satisfactorily with regard to digital documents containing alphabet characters, but they still have a long way to go as far as formulae and graphs are concerned. In general, the most spread digital documents are in PDF format. However, in the case of mathematical contents, they are not accessible at all, since formulae are usually unreadable by screen readers. Currently, a standard and fast method for inserting accessible formulae into a PDF document is still lacking despite it is a very important issue for spreading accessible digital scientific documents. In this paper, we propose a method for automatically generating a PDF document with mathematical contents accessible by assistive technologies for visually impaired people. Specifically, we have developed a LaTeX package that produces a final PDF document where the formulae are totally accessible by screen readers and braille displays.

Keywords: accessibility of formulae, assistive technology, LaTeX, pdf document, visually impaired people

1 Introduction

In this paper, we propose a method for automatically generating a PDF document with formulae accessible by assistive technologies for visually impaired people. Assistive technologies (screen readers and braille displays) perform satisfactorily with regard to digital documents containing alphabet characters, but they still have a long way to go as far as formulae and graphs are concerned. A comprehensive overview about this problem can be found in [1] and [2].

Many studies have been conducted in order to improve the accessibility of digital documents with mathematical contents. For instance, MathPlayer ensures accessibility of formulae inserted by using MathType in Word documents [10]. Another way for creating accessible mathematical documents is given by the MathML language (see [4] for further information). However, accessibility of

such documents is heavily affected by the versions of browsers, operative systems and screen readers, making this solution very unstable. A system used by blind people for reading and writing mathematics is the LAMBDA system (Linear Access to Mathematics for braille Device and Audio-synthesis). Mathematical language in LAMBDA is designed so that every symbol can be directly translated into words. For further details on LAMBDA we refer to [5]. Unfortunately, this system does not help to spread accessible digital documents, since it is only used by visually impaired people and it is not a standard for the realization of documents by sighted people. A very used language for producing scientific documents is the LaTeX language, which is a standard in the scientific community. Assistive technologies can directly manage LaTeX documents. In this case, visually impaired people need to learn LaTeX in order to understand the commands. However, there are software which facilitate LaTeX comprehension and usability; one of them is BlindMath [9]. Moreover, some converters from LaTeX to braille exist, see, e.g., [8] and [3].

In general, the most spread digital documents are in PDF format. However, in the case of mathematical contents, they are not accessible at all, since formulae are usually unreadable by screen readers because they are bidimensional as images. None of the above systems allows to directly produce accessible formulae in PDF documents. This could be possible only performing specific tasks. For instance, using the Word editor, if each formula is manually tagged by the author (by using the alternative text), such a comment will be kept when the corresponding PDF file will be generated and it will be read by the screen reader. However, this procedure does not help to improve the presence of accessible PDF documents, since it is a very boring and time consuming method. It is very hard to think that an author performs these actions for the realization, e.g., of a book. Currently, a standard and fast method for inserting accessible formulae into a PDF documents is still lacking despite it is a very important issue for spreading accessible digital scientific documents. In [11] standard guidelines for accessibility of PDF documents are presented. Moreover, in [6] and [7], an overview about accessibility of PDF documents is provided with a focus on mathematical contents. In these works, the author also proposes possible solutions and guidelines for producing accessible formulae by embedding LaTeX and MathML codes. However, a working and definitive method is not provided.

In this paper, we show a LaTeX based method for an automatized production of accessible PDF documents with mathematical contents. Specifically, we have developed a LaTeX package that produces a final PDF document where the formulae are totally accessible by screen readers and braille displays.

2 Description of the LaTeX package ‘accformulae’

When a PDF document is generated starting from LaTeX, formulae are not accessible by screen readers and braille displays. They can be made accessible by inserting a hidden comment, i.e., an actual text, similarly to the case of web pages or Word documents. This can be made, e.g., by using the LaTeX package

pdfcomment.sty or using an editor for PDF files like Adobe Acrobat Pro. In any case, this task must be manually performed by the author and it is surely inefficient, since the author should write the formulae and, in addition, insert a description for each formula. Note also that the package pdfcomment.sty does not allow to insert special characters like ‘backslahs’, ‘brace’, etc, in the comment. Moreover, with these solutions, the reading is bothered since the screen reader reads incorrectly the formula and then the correct comment of the formula. In Figure 1, we show the LaTeX code for generating a PDF document containing a simple formula with a comment manually inserted. When the screen reader accesses the PDF document, the formula will be read ‘square root 1 plus 5 2 begin fraction numerator 1 plus square root of 5 over 2 end fraction’, i.e., before reading the correct comment ‘begin fraction numerator 1 plus square root of 5 over 2 end fraction’ the screen reader reads incorrectly the formula ‘square root 1 plus 5 2’.

```

\documentclass[a4paper,11pt]{article}

\usepackage{pdfcomment}

\begin{document}

A simple formula:

\begin{equation}
\pdfcomment{\frac{1 + \sqrt{5}}{2}}
{begin fraction numerator 1 +
square root of 5 over 2 end fraction}
\end{equation}

\end{document}

```

A simple formula:

$$\frac{1 + \sqrt{5}}{2}$$

begin fraction numerator 1 + square root of 5 over 2 end fraction

Fig. 1: LaTeX commands for generating a simple formula in a PDF document using the package pdfcomment.sty and the corresponding PDF output

There are also some LaTeX packages that try to improve the accessibility of PDF documents produced by LaTeX. In particular the packages accsupp.sty

(available at <https://ctan.org/pkg/accsupp>) and `accessibility_meta.sty` (available at <https://github.com/AndyClifton/AccessibleMetaClass>) has been developed in order to obtain tagged PDF documents. However, both packages do not solve the problem of the accessibility of formulae. The package `accsupp.sty` develops some interesting tools for commenting formulae using also special characters (possibility that is not available in the `pdfcomment.sty` package). Moreover, this is not an automatized method, since the comment must be manually inserted by the author. The package `accessibility_meta.sty` is an improved version of the package `accessibility.sty`. This package allows the possibility of inserting several tags for sections, links, figures and tables. However, even if these tags are recognized by the tool for checking tags of Acrobat Reader Pro, they are not always recognized by the screen readers. Moreover, this package does not manage formulae and it is not uploaded in the official CTAN repository (some errors may appear when compiling a LaTeX file using this package).

Our package, named `accformulae.sty`, solves completely this problem, since it is able to automatically produce an actual text corresponding to the LaTeX commands that generate the formulae. This actual text is hidden in the PDF document but the screen reader reads it without reading any incorrect sequence before.

The package `accformulae.sty` uses the command `\BeginAccSupp` defined in the existing package `accsupp.sty`. Such a command has been modified in order to obtain an actual text readable by the screen reader.

We have treated the most used environments for inserting formulae, i.e., `equation`, `equation*`, `\[`, `$$`, `\(`, `$`. Hence, any formula inserted using one of these environments is accessible in the corresponding PDF document.

In Figure 2, we show the LaTeX code for generating the same PDF document shown in Figure 1. We can observe that, in this case, the author has to write the formula without adding anything else. In Figure 2, we also show some lines of the source code of the PDF file, where there is the actual text automatically generated by our package. The screen reader will read correctly the LaTeX command `\frac {1+ \sqrt {5}}{2}`. Moreover, we have created a JAWS dictionary that provides the reading in the natural language in the case that the user does not know the LaTeX commands.

A video that shows the performances of the screen readers NVDA and JAWS on a sample PDF document generated by our package is available at <http://www.integr-abile.unito.it/demo-accformulae.php>.

3 User involvement

The package has been evaluated by four visually impaired people, named subjects A, B, C and D. They are all blind, but it is noticed that, while two of them are blind from birth and know the braille code, the others became blind during their life and they do not use the braille code. Thus, during the tests two used a refreshable braille display and two did not. Subjects A and C used the screen reader NVDA, whereas subjects B and C used JAWS. Moreover, subjects A

```

\documentclass[a4paper,11pt]{article}

\usepackage{accformulae}

\begin{document}

A simple formula:
\begin{equation}
\frac{1 + \sqrt{5}}{2}
\end{equation}

\end{document}

stream
BT
/F15 10.9091 Tf 134.765 704.247
Td [(A)-333(simple)-334(form)28(ula:)]TJ
ET
/S/Space
<</ActualText(\040\frac\040{1\040+\040\sqrt\040{5}}{2}\040)>>
BT
/F15 10.9091 Tf 280.489 691.232
Td [(1)-222(+)]TJ/F21 10.9091 Tf 18.788 9.024 Td [(p)]TJ
ET
q
1 0 0 1 308.368 700.475 cm
[ ]0 d 0 J 0.436 w 0 0 m 5.455 0 1 s
Q
BT
/F15 10.9091 Tf 308.368 691.232 Td [(5)]TJ
ET
q
1 0 0 1 280.489 686.579 cm
[ ]0 d 0 J 0.436 w 0 0 m 33.333 0 1 s
Q
BT
/F15 10.9091 Tf 294.428 676.369 Td [(2)]TJ
ET

```

Fig. 2: LaTeX commands for generating a simple formula in a PDF document using the package `accformulae.sty` and the corresponding source code of the PDF output (note that code `\040` stands for the space in the PDF source)

and B have a very good knowledge of the LaTeX language, subject B is an intermediate user, while subject D is a beginner. Some information about the participants are summarized in Table 1.

Table 1: Participants to the evaluation

Subject	Screen Reader	Braille Display	LaTeX User
A	NVDA	Yes	Expert
B	JAWS	Yes	Expert
C	NVDA	No	Intermediate
D	JAWS	No	Beginner

All the participants tested some PDF documents produced by using our package. All of them have been able to read correctly the formulae inside the document; indeed, they observed that, when a formula is encountered, the screen reader reads the corresponding LaTeX commands. They also noticed that in the PDF document the formula is not marked by dollars (or by the used environment) as in a LaTeX document. Participants A and B reported that in the braille display appears what the screen reader reads. The heading levels are still lacking in the PDF documents.

The PDF documents have been accessed by Acrobat Reader DC, since it is the most used PDF viewer and it is well integrated with screen readers. If different PDF viewers are used, it is possible that the formulae are not read correctly.

Participants B and D also tested the JAWS dictionary that we have realized. They have been able to read correctly the formulae, with the difference that in this case they are read in natural language. Participant B reported that in the braille display it is still written the corresponding LaTeX commands of the formulae.

Finally, participants A and B successfully used the package for generating themselves an accessible PDF document containing formulae.

4 Conclusion

We have developed a LaTeX package that automatically generates comments to formulae when the PDF document is produced by LaTeX. The comments are hidden in the PDF document and they contain the LaTeX commands that generate the formulae. In this way, an accessible PDF document containing formulae is generated. Indeed, screen readers are able to access the comment when processing a formula and reading it. Moreover, we have created a JAWS dictionary that provides the reading in the natural language in the case that the user does not know the LaTeX commands.

As future work, we intend to release the package `accformulae.sty` on the official CTAN repository. In the final version of the package, we will add all the

environments used for inserting formulae (for instance, in the current version of the package, the environment ‘align’ is not managed). Another issue, which we would like to address, concerns the heading levels. Currently a PDF file produced by LaTeX does not contain heading levels (and the existing packages, like accessibility.sty, do not fix this problem). Finally, we aim at developing a LaTeX package that produces PDF/UA documents.

5 Acknowledgements

The authors wish to thank the bank foundation ‘Fondazione Cassa di Risparmio di Torino’, LeoClub (Biella, Italy) and the several volunteers with visual impairment who provided their fundamental contribution.

References

1. Archambault, D., Stoger, B., Fitzpatrick, D., Miesenberger, K.: Access to scientific content by visually impaired people. Upgrade 2, 14 pages (2007).
2. Armano, T., Capietto, A., Illengo, M., Murru, N., Rossini, R.: An overview on ICT for the accessibility of scientific texts by visually impaired students. Proceedings Conference SIREM-SIE-L 2014, Perugia (Italy), 119–122 (2014).
3. Batusic, M., Miesenberger, K., Stoger, B.: LaBraDoor, a contribution to making mathematics accessible for the blind, Proceedings of 6th International Conference on Computers Helping People with Special Needs, ICCHP 1998, Oldenbourg, Wien, Munchen (1998).
4. Bernareggi, C., Archambault, D.: Mathematics on the web: emerging opportunities for visually impaired people. Proceeding W4A '07 Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), Banff (Canada), 108–111 (2007).
5. Bernareggi, C.: Non-sequential mathematical notations in the LAMBDA system. Computers Helping People with Special Needs, Lecture Notes in Computer Science 6180, 389–395 (2010).
6. Moore, R.: Ongoing efforts to generate tagged PDF using pdfTEX. TUGboat 30, 170–175 (2009).
7. Moore, R.: PDF/A-3u as an Archival Format for Accessible Mathematics. Watt, Stephen M., Davenport, James H., Sexton, Alan P., Sojka, Petr, Urban, Josef (eds.) CICM, Lecture Notes in Computer Science 8543, 184–199 (2014).
8. Papasalouros, A., Tsolomitis, A.: A direct TeX-to-Braille transcribing method. Journal of Science Education for Students with Disabilities 20, Article 5 (2017).
9. Pepino, A., Freda, C., Ferraro, F., Pagliara, S., Zanfardino, F.: ‘BlindMath’ a New Scientific Editor for Blind Students. Computers Helping People with Special Needs, Lecture Notes in Computer Science 4061, 1171–1174 (2006).
10. Soiffer, N.: Mathplayer: web-based math accessibility. Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility, New York (USA), 204–205 (2005).
11. Uebelbacher, A., Bianchetti, R., Riesch, M.: PDF Accessibility Checker (PAC 2): The First Tool to Test PDF Documents for PDF/UA Compliance. Computers Helping People with Special Needs, Lecture Notes in Computer Science 8547, 197–201 (2014).