An overview on ICT for the accessibility of scientific texts by visually impaired students

Tiziana ARMANO[1], Anna CAPIETTO[1], Marco ILLENGO[1], Nadir MURRU[1], Rosaria ROSSINI[2]

[1] *University of Turin, Department of Mathematics "G. Peano",*

[2] *University of Turin, Department of Computer Science,*

tiziana.armano@unito.it, anna.capietto@unito.it, illengo@gmail.com, nadir.murru@unito.it, rossini@di.unito.it

Abstract

We provide a brief overview on the most common systems used to deal with university teaching materials (UTM) with mathematical contents, highlighting their problems in terms of accessibility. We shall describe some instruments and methods that universities could offer to visually impaired students for reading UTM and we shall point out software enabling visually impaired students and their teachers to write texts with mathematical contents.

Keywords: mathematics accessibility, web accessibility, visually impaired

Introduction

People with sensorial disabilities willing to undertake University courses have nowadays new instruments at their disposal.

Visually impaired people strongly benefit from speech synthesis software such as NVDA or JAWS (Windows), ORCA (Linux), and VoiceOver (Apple). Blind people also use braille displays, whereas partially sighted people may use any combination of screen magnifiers and color customization (their requirements may vary, due to high diversity in partial eyesight).

Assistive technologies perform satisfactorily with regard to texts, but they still have a long way to go as far as formulae and graphs are concerned. Indeed, these are usually represented in two dimensions, while language follows a one-dimensional construction.

At the moment, the content of a graph can be either captioned or embossed. Captions can only provide a limited description; embossed paper may convey a better feeling of the whole picture, although it requires a dedicated printer (braille embosser), larger and thicker paper sheets, and possibly several prints for each graph. In particular, it is possible to print both braille and ink on the same page, thus allowing for a substantial interaction between sighted and blind people.

Formulae can be written in 8-dots braille, although there is no international standard: several countries have developed their own braille code for mathematics and a serious work in the direction of a unified braille system is far from being accomplished. Speech synthesis allows a computer to read the transcription of any formula, although the resulting text may become impractically long (e.g. ½ might be read as "begin fraction numerator one denominator two end fraction"). Such transcriptions can be readily obtained when the software itself is made aware of the mathematical structure, that is, when formulae are written by means of a *markup* language: a computer language interspersed with special annotations marking the beginning and the end of any mathematical structure (subscripts, superscripts, fractions, and so on). We refer to the work of Archambault et al. (2007) for a good survey on this topic.

Some standard instruments

Availability of a large amount of textual information, such as slides, books, lecture notes, scientific articles, on the web or in university digital libraries helps students in their educational path from elementary school to university. In this section we give an overview of the most common instruments that students and professors use to create scientific documents.

Microsoft Word is the most widespread text editor. It has a graphical interface that allows to write text and to insert formulae and images. However, graphs and formulae inserted in Word documents are generally not accessible.

HTML and XHTML are markup languages which are used for creating web pages and are helpful to make available UTM to a wide audience. As mentioned above, markup languages are suitable to be managed by screen readers and Braille displays. However, graphs and formulae are generally inserted as images, which are clearly not accessible.

LaTeX is a markup language widely used by the scientific community for the production of high quality documents with mathematical contents. Screen readers and braille displays can directly access raw LaTeX documents containing formulae. However, even once visually impaired students have learned this language, the output provided by screen readers and braille displays is not completely handy; for instance, the fraction ½ is represented by the LaTeX command "\frac{1}{2}".

We also remark that most UTM is exported and supplied in *.pdf*.

On accessible reading

We describe which kinds of UTM are accessible to visually impaired students, as well as some software and methods that can be used in this field.

In the case of Word documents, there is no way to make graphs accessible, while MathPlayer ensures accessibility of formulae inserted by using MathType. See Bernareggi and Archambault (2007) for a comprehensive about MathPlayer.

HTML and XHTML documents are often obtained starting from documents written with word processors. In these documents, images can have a hidden caption (the "alt" attribute) that is accessible. Moreover, formulae can be written in MathML embedded in these documents. In this case, formulae are accessible on Apple machines using VoiceOver on Safari browser or in the Windows framework using JAWS/NVDA and MathPlayer combined with Internet Explorer. Unfortunately, MathPlayer is only compatible with Internet Explorer 9 (or older versions) which is no longer supported by recent operative systems.

Speech and braille assistive technologies directly read LaTeX documents. In this case, visually impaired students need to learn LaTeX in order to understand the commands. However, there are several software which facilitate LaTeX comprehension and usability; one of them is BlindMath. The project LaTeX-access provides a translation of LaTeX commands into statements easier to understand; as a result, the usability by visually impaired people is improved. Moreover, text editors (e.g. Kile) often offer color customization of LaTeX commands, thus helping partially sighted people as well. Finally, we remark that several programs translate LaTeX documents into HTML and XHTML documents, where formulae are inserted with MathML and are therefore accessible.

In Uebelbacher et al. (2014) standard guidelines for accessibility of *.pdf* documents are presented, as well as the open source software PDF Accessibility Checker 2 (PAC2), which can evaluate the accessibility of any *.pdf* document. In many cases *.pdf* files are read by screen readers; however it is frequent (especially in case of scanned documents) that such files are not accessible at all. In these cases an Optical Character Recognition (OCR) software is required in order to obtain an editable file.

Recently, some systems specifically designed for blind people have been developed. The Infty project, developed by Suzuki et al. (2004), works on a LaTeX-based language and has its own speech synthesis software (ChattyInfty) that reads formulae written in Infty documents. Moreover, Infty Reader is an OCR for the automatic recognition of formulae, although the automatic recognition is not optimal and often requires corrections, especially if the starting document is not in English. Infty Reader is the only OCR that performs an automatic recognition of complex formulae (e.g. integrals and roots). Indeed, existing OCRs (like Abbyy Fine Reader and Tesseract) have optimal performances in the recognition of normal text but, up to now, they can only recognize very simple formulae. Furthermore, apps like MathPad (for Apple) are able to recognize single handwritten formulae, but they do not perform automatic recognition for a complete document.

A system widely used by blind people is the LAMBDA system (Linear Access to Mathematics for Braille Device and Audio-synthesis). LAMBDA is composed of a markup language, a dedicated editor and a MathML converter. Mathematical language in LAMBDA is designed so that every symbol can be directly translated into words. This translation is implemented on the fly in the editor, which can therefore be used also by visually impaired people. The converter imports and exports XHTML documents containing MathML code. For further details on LAMBDA we refer to Bernareggi (2010).

On accessible writing

In this section, we review the above instruments, pointing out how they can be used by visually impaired people for writing texts with mathematical contents.

In general, Word is not a good solution for writing scientific texts, although the LeanMath editor developed by Gardner (2014) allows visually impaired people to access to the graphical interface of MathType through speech synthesis and hotkeys.

HTML and XHTML together with MathML are very hard to learn and are not handy for neither sighted nor visually impaired people. XHTML should be preferred to HTML, since its documents are more accessible and can easily be imported by specific systems like LAMBDA.

LaTeX is well-known by professors of scientific courses; students approaching scientific degrees are highly suggested to learn it. Editors for LaTeX documents are usually accessible; moreover software like BlindMath and LaTeX-access make the use of TeX easier. Furthermore, any LaTeX document can be imported by Infty or can be converted into an XHTML document which, in turn, can be imported by LAMBDA.

Infty Editor interacts well with LaTeX documents, but its graphical interface is not accessible. Thus, it could be recommended to sighted people in order to produce accessible documents, but this is not a feasible solution for visually impaired students.

Finally, in LAMBDA, the mathematical editor is especially designed for blind people and it is accessible using hotkeys, braille displays, and speech synthesis. Furthermore, this editor offers support for partially sighted people by using a colour code. Although LAMBDA is a good solution for visually impaired people willing to write texts with formulae, it is not completely suitable at university level, since the list of mathematical symbols is neither complete nor adaptable.

Future work

We intend to focus on an in-depth analysis and a systematization of the aforementioned topics. Moreover, we will concentrate on the accessibility of UTM via mobile devices and on the use of 3D printers. We shall also focus on developing a new OCR for formulae, as well as on realising the Italian version of the Infty project.

References

Archambault, D., Stöger, B., Fitzpatrick, D., & Miesenberger, K., (2007). *Access to scientific content by visually impaired people.* Upgrade, VIII(2), 14 pages.

Bernareggi, C., (2010). *Non-sequential Mathematical Notations in the LAMBDA System*. Computers Helping People with Special Needs, Lecture Notes in Computer Science, 6180, 389-395.

Bernareggi, C., & Archambault, D., (2007). *Mathematics on the web: emerging opportunities for visually impaired people.* Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), 108-111.

Gardner, J. A., (2014). *The LEAN Math Accessible MathML Editor.* Computers Helping People with Special Needs, Lecture Notes in Computer Science, 8547, 580-588.

Uebelbacher, A., Bianchetti, R., & Riesch, M. (2014). *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.

Suzuki, M., Kanahori, T., Ohtake, N., & Yamaguchi, K., (2004). *An Integrated OCR Software for Mathematical Documents and Its Output with Accessibility*, Computers Helping People with Special Needs, Lecture Notes in Computer Science, 3118, 648-655.

Acknowledgments

This work has been developed in the framework of an agreement between IRIFOR/UICI (Institute for Research, Education and Rehabilitation/Italian Union for the Blind and Partially Sighted) and Turin University.