An Automatized Method Based on LaTeX for the Realization of Accessible PDF Documents Containing Formulae

Anna Capietto and Sandro Coriasco
Dipartimento di Matematica "G. Peano"
Università degli Studi di Torino

TUG Meeting 2018

Rio de Janeiro, 20/07/2018

An Automatized Method Based on LaTeX for the Realization of Accessible PDF Documents Containing Formulae

D. Ahmetovic, T. Armano, C. Bernareggi, M. Berra,
A. Capietto, S. Coriasco, N. Murru, A. Ruighi, E. Taranto
Dipartimento di Matematica "G. Peano"
Università degli Studi di Torino

TUG Meeting 2018

Rio de Janeiro, 20/07/2018

Introduction

As an introduction, we first describe some of the activities of the

Laboratory for the research and experimentation of new assistive technologies for STEM "S. Polin"

It was established within the framework of the project "For an accessible and inclusive mathematics" of the Department of Mathematics "G. Peano", Turin University, which is the technological/scientific part of the Turin University project "Disability and New Technologies", coordinated by Prof. M. Pavone (professor of Special Education in the Department of Philosophy and Educational Sciences of the University of Turin and Deputy Rector for the Disability), and by Prof. A. Capietto (referee for students with special educational needs at the UniTO Math. Dept.), aimed at the optimization, testing and development of new technologies for the access to advanced studies of students with physical or sensorial disabilities.

Main supporters and partners

- University of Torino;
- Department of Mathematics “G. Peano”, University of Torino;
- Foundation “Cassa di Risparmio di Torino”;
- Foundation Specchio dei Tempi (“La Stampa”, Torino);
- I.Ri.Fo.R./UICI - Institute for Research, Training and Rehabilitation/Italian Union of the Blind and Visually Impaired people.

Our activities

- Research;
- Experimentation of assistive technologies;
- Dissemination on the territory ("Terza Missione").

We concentrate on the digitalization/computerized-processing of STEM (Science Technology, Engineering and Math) contents, their applications and accessibility. Special attention is devoted to visual impairment, in the framework of an agreement between Turin University and I.Ri.Fo.R./UICI.

The group

- A. Capietto (Full professor at the Dept. of Mathematics, UniTO);
- S. Coriasco (Associate professor at the Dept. of Mathematics, UniTO);
- D. Ahmetovic ('assegnista di ricerca' at UniTO);
- G. Airò Farulla (Università Cà Foscari, Venice);
- T. Armano (research technician at the Dept. of Mathematics, UniTO);
- C. Bernareggi ('borsista di ricerca' at the Dept. of Mathematics, UniTO, and UniMI);
- M. Berra ('assegnista di ricerca' at the Dept. of Mathematics, UniTO);
- M. Borsero ('borsista di ricerca' at the Dept. of Mathematics, UniTO, and teacher at secondary school);
- M. Bracco (teacher at secondary school);

- S. Kobal (law student at the University of Trento, volunteer);
- A. Mazzei (researcher at the Computer Science Dept., UniTO);
- N. Murru ('borsista di ricerca' at the Dept. of Mathematics, UniTO);
- A. Panzarea (I.Ri.Fo.R./UICI, volunteer);
- R. Ricci (student at the Politecnico di Torino, volunteer);
- R. Rossini (Ph.D. in Computer Science, volunteer);
- A. Ruighi (graduate/Ph.D. student at PoliTO- UniTO);
- E. Taranto (University of Catania);
- E. Tornavacca (Società Reale Mutua Assicurazioni, Torino).

We are a member of AsTech/CINI ("Consorzio Italiano Nazionale Informatica").

D.A.P.A.R.I.

D.A.P.A.R.I. stands for Disabilità in Azienda, Professionalità Avanzata, Ricerca, Inclusione.
A project with Società Reale Mutua Assicurazioni (Torino) and the Dipartimento di Scienze economico-sociali e matematico-statistiche of the University of Torino.
D.A.P.A.R.I. aims at supporting persons with disabilities (in particular visually impaired persons) from university training to the world of work.
Starting from November 2014, Dott. Tornavacca (a person with low vision), graduate in Mathematics and working for Reale Mutua Assicurazioni, works during 40 days (in a year) at the Dept. of Mathematics.

Improving the occupability of persons with visual impairment

In cooperation with Città Metropolitana of Torino (Fondo Regionale Disabili), I.Ri.Fo.R. and D.A.P.A.R.I.

1. A 14-hour course for visually impaired people in order to reinforce their IT competence addressed to the world of work;
2. 9 payed internships, for three months, where the partecipants developed the competences acquired during the course;
3. 2 internships have turned in definite positions.

An opportunity for persons with visual impairment to widen their work opportunities, and for the world of work to get informed of the possibility of having a person with visual impairment as an effective corporate resource.

Visual impairment

People with visual disabilities willing to undertake University courses have nowadays new instruments available, such as:

- Screen readers;
- Braille displays;
- Screen magnifiers;
- Speech synthesis software;
- ...

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, the latter are usually represented in two dimensions, while language follows a one-dimensional construction.

LaTeX packages and PDF files accessibility

Fact: A great deal of teaching material produced within mathematical courses (and most of the mathematical research papers/books ...), very often consists of PDF files produced with LaTeX.

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, that is, an actual text (similarly to the case of web pages).

There have been many efforts to achieve accessible PDF documents through the implementation of LaTeX packages, like, e.g., pdfcomment and accsupp, and, recently, the last updates of pdfx. However, with them the "tagging/commenting procedure", as far as we understood, should be manually performed by the author (for instance, the author should write the formulae and, in addition, insert a description for each formula ...). Note also that, for instance, the package pdfcomment does not allow to insert special characters like 'backslash', 'brace', etc, in the comment.

LaTeX packages and PDF files accessibility

Notice that, with some of the previously mentioned solutions, the reading is difficult, since the screen reader reads incorrectly the formula, and then the correct comment of the formula. We aimed at:

- making mathematical formulae in PDF files generated by LaTeX correctly accessible to readers/Braille displays;
- avoiding, as far as possible, that the author needs too many extra efforts/to insert additional LaTeX commands to achieve this goal.

For simplicity, given our aims, we started from the package accsupp, which also develops useful tools for commenting formulae (allowing also special characters in the comment). Again, by itself, it would not provide this automatically, since the comment should be manually inserted by the author.

The axessibility LaTeX package

Our package automatically produces 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. This is obtained by marking with suitable tags both the actual text as well as the formula.

The axessibility LaTeX package – Implementation

```
...
\renewcommand*{\BeginAccSupp}[1]{%
  \begingroup
   ...
    \edef\ACCSUPP@span{%
      /S/Span<<%
        \ifx\ACCSUPP@Lang\relax
        \else
          /Lang\ACCSUPP@Lang
        \fi
        ...
      >>%
    }%
    \ACCSUPP@bdc
    \ACCSUPP@space
  \endgroup
}
```

The axessibility LaTeX package – Implementation

```
...
\long\def\wrap#1{
\BeginAccSupp{method=escape,
                ActualText=\detokenize\expandafter{#1}}
 #1
\EndAccSupp{}%
}
...
\renewenvironment{equation}{%
 \incr@eqnum
   \mathdisplay@push
   \st@rredfalse \global\@eqnswtrue
  \mathdisplay{equation}%
   \collect@body\wrap\auxiliaryspace}{%
   \endmathdisplay{equation}%
  \mathdisplay@pop
  \ignorespacesafterend}...
```

An example – Typeset PDF

The golden mean

The golden mean is the number

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

that is, the root larger in modulus of

$$x^2 - x - 1. \tag{1}$$

It is usually defined as the ratio of two lengths a and b such that

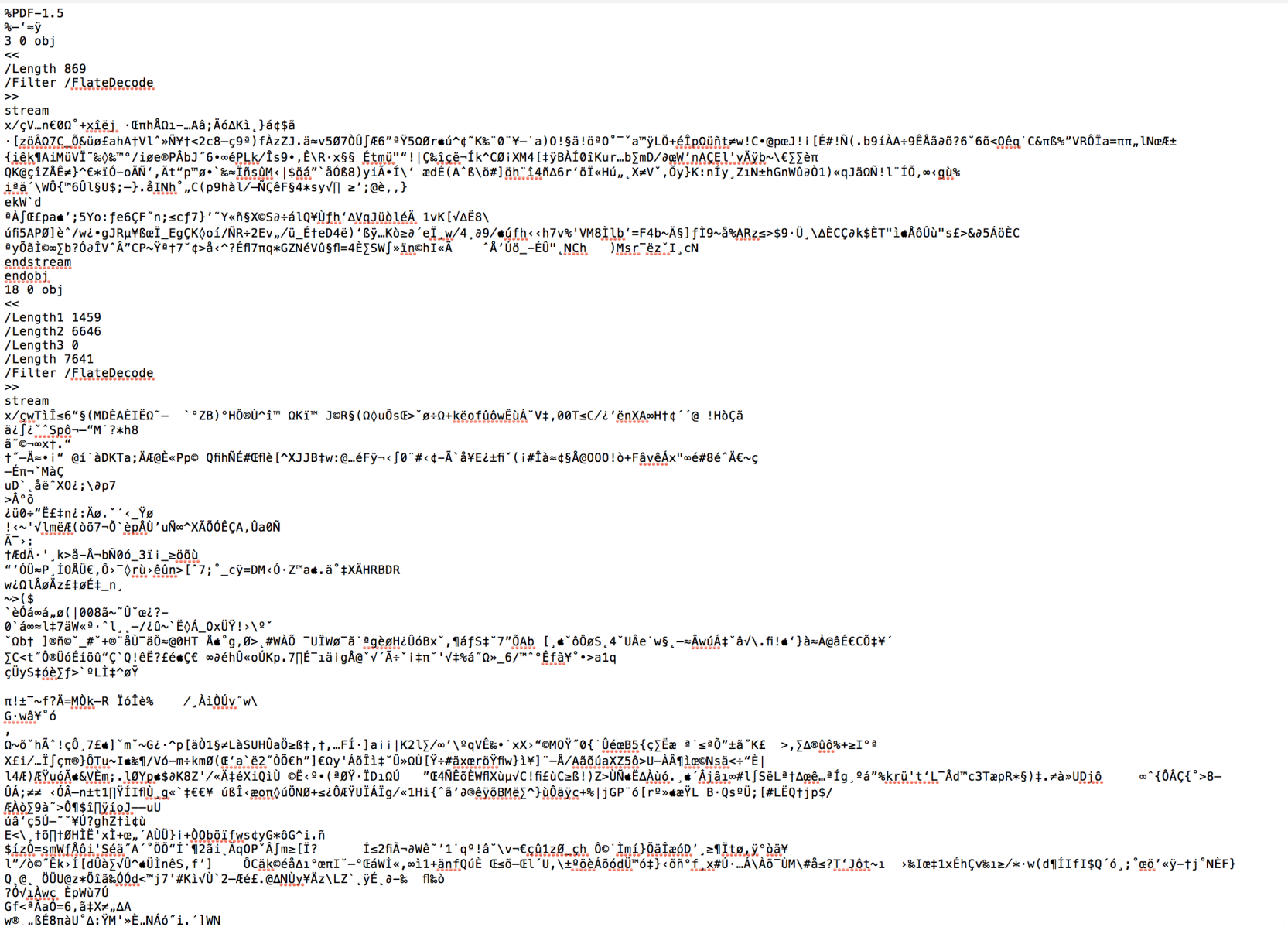
$$(a+b) : a = a : b.$$

Let x be the ratio $\frac{a}{b}$, we have $\frac{a+b}{a} = 1 + \frac{1}{x}$, from which we get the equation $x^2 = x + 1$.

An example – LaTeX code

```
\documentclass[a4paper,11pt]{article}
\usepackage{amsmath}%\usepackage{axessibility}
\title{The golden mean}\author{}\date{}
\begin{document}
\maketitle
The golden mean is the number
\[\frac{1 + \sqrt{5}}{2},\]
that is the root larger in modulus of
\begin{equation} x^2 - x - 1. \end{equation}
It is usually defined as the ratio of two lengths
 \(a\) and \(b\) such that
\begin{equation*} (a+b) : a = a : b. \end{equation*}
Let \(x\) be the ratio \( \frac{a}{b} \), we have
\( \frac{a+b}{a} = 1 + \frac{1}{x} \), from which
we get the equation \(x^2 = x + 1\).
\end{document}
```

An example – Typeset PDF content without axessibility



An example – Typeset PDF content with axessibility

```
%PDF-1.6
%–‘≈ÿ
3 0 obj
<<
/Length 3497
>>
stream
BT
/F16 17.2154 Tf 237.491 660.4 Td [(The)-302(golden)-302(mean)]TJ/F15 10.9091 Tf -102.726 -72.179 Td [(The)-333(golden)-334(mean)-333(is)-333(the)-334(n)28(um)28(b)-28(er)]TJ
ET
/S/Span<</ActualText(\040\\frac\040{1\040+\040\\sqrt\040{5}}{2},)>>BDC
BT
/F15 10.9091 Tf 278.974 564.248 Td [(1)-222(+)]TJ/F36 10.9091 Tf 18.788 9.024 Td [(p)]TJ
ET
q
1 0 0 1 306.853 573.49 cm
[]0 d 0 J 0.436 w 0 0 m 5.455 0 l S
Q
BT
/F15 10.9091 Tf 306.853 564.248 Td [(5)]TJ
ET
q
1 0 0 1 278.974 559.595 cm
[]0 d 0 J 0.436 w 0 0 m 33.333 0 l S
Q
BT
/F15 10.9091 Tf 292.913 549.384 Td [(2)]TJ/F35 10.9091 Tf 20.59 7.484 Td [(;)]TJ
ET
EMC
BT
/F15 10.9091 Tf 117.828 529.854 Td [(that)-333(is)-334(the)-333(ro)-28(ot)-333(larger)-333(in)-334(mo)-27(dulus)-334(of)]TJ
ET
/S/Span<</ActualText(\040\040x^2\040-\040x\040-\0401.\040)>>BDC
BT
/F35 10.9091 Tf 270.979 505.345 Td [(x)]TJ/F18 7.9701 Tf 6.235 4.505 Td [(2)]TJ/F36 10.9091 Tf 7.156 -4.505 Td [(\000)]TJ/F35 10.9091 Tf 10.909 0 Td [(x)]TJ/F36 10.9091 Tf
8.659 0 Td [(\000)]TJ/F15 10.9091 Tf 10.909 0 Td [(1)]TJ/F35 10.9091 Tf 5.455 0 Td [(:)]TJ
ET
EMC
BT
/F15 10.9091 Tf 462.544 505.345 Td [(\0501\051)]TJ -344.716 -24.508 Td
[(It)-333(is)-334(usually)-333(de\014ned)-333(as)-334(the)-333(ratio)-333(of)-333(t)27(w)28(o)-333(lengths)]TJ
ET
/S/Span<</ActualText(a)>>BDC
BT
/F35 10.9091 Tf 344.465 480.837 Td [(a)]TJ
ET
EMC
BT
/F15 10.9091 Tf 353.867 480.837 Td [(and)]TJ
ET
/S/Span<</ActualText(b)>>BDC
BT
/F35 10.9091 Tf 375.08 480.837 Td [(b)]TJ
ET
EMC
BT
/F15 10.9091 Tf 383.398 480.837 Td [(suc)28(h)-334(th)1(at)]TJ
ET
/S/Span<</ActualText(\040\040\(a+b\)\040:\040a\040=\040a\040:\040b.\040)>>BDC
BT
/F15 10.9091 Tf 255.036 456.329 Td [(\050)]TJ/F35 10.9091 Tf 4.243 0 Td [(a)]TJ/F15 10.9091 Tf 8.191 0 Td [(+)]TJ/F35 10.9091 Tf 10.909 0 Td [(b)]TJ/F15 10.9091 Tf 4.681 0 Td
[(\051)-278(:)]TJ/F35 10.9091 Tf 13.334 0 Td [(a)]TJ/F15 10.9091 Tf 8.796 0 Td [(=)]TJ/F35 10.9091 Tf 11.515 0 Td [(a)]TJ/F15 10.9091 Tf 8.797 0 Td [(:)]TJ/F35 10.9091 Tf
6.061 0 Td [(b:)]TJ
ET
```

An example – VoiceOver

Try reading with Acrobat Reader DC & VoiceOver

`http://www.integr-abile.unito.it/share/VoiceOver.mp4`

Try reading with NVDA+JAWS

`http://www.integr-abile.unito.it/share/NVDAandJAWS.mp4`

The axessibility LaTeX package – Complementary scripts

In addition to the package, we also provide scripts that complement package functionalities.

- Preprocessing scripts. It is possible, in principle, to apply our package to an already existing LaTeX document. In this case, it is necessary to preprocess the document in order to replace some of the unsupported commands and environments. We provide a preprocessing script to handle some of these cases at our Github repository.
- Expansion of user macros. Note that custom macros used by the author within the formulae are copied as-is into the actual text in the hidden comment. This macros may bear no meaning for other readers, so it may be more meaningful to expand those macros into the original LaTeX commands. We provide a script that can parse LaTeX document and replace all the user macros within the formulae with their expanded definitions at our Github repository.

The axessibility LaTeX package – Additional resources

- Screen reader dictionaries. LaTeX commands that are included as actual text in the hidden comments corresponding to formulae may appear awkward when read by the screen reader. We provide dictionaries for JAWS and NVDA screen readers that convert LaTeX commands into natural language. Please note that the Braille refreshable display will still show the formulae in their original LaTeX representations. The dictionaries will be available at our Github repository.

Research – OCR

In order to obtain an accessible text it is necessary to start from an editable format, in order to, if necessary, modify it in a suitable way.
An OCR sw trasforms a non-editable text (e.g., .pdf, .jpg, etc...) into an editable one. To do this, two main steps are performed:

1. An image segmentation algorithm
2. A character recognition algorithm

Research – OCR

We are developing the first European OCR prototype that automatically recognizes both text and formulae.

In this case, an OCR must recognize a very large quantity of patterns and existing pattern recognition algorithms have troubles to converge and perform recognition.

In presence of formulae, image segmentation is an hard task as well, since cutting positions can occur vertically, horizontally and diagonally (whereas, printed touching characters can be separated by vertical cuts).

At our best knowledge, at present, the only OCR of this kind (Infty Project) is developed by a network of Japanese universities. Authors implement completely different techniques.

Research - Other disabilities

1. Production of a database of the videotapes of the classes of the Bachelor's Degree in Mathematics.

2. For persons with motor disabilities (upper limbs), we study and optimize the access to the computer by voice commands and the automatic transcription of vocal notes (MathTalk).

Experimentation of assistive technologies

- Accessibility of Moodle 2;
- Accessibility of University admission tests;
- Consultation for other universities Disability Offices;
- Improved accessibility of PDF-LaTeX files;
- Accessibility of graphs.

Dissemination on the territory ("Terza missione")

- Unito - I.Ri.Fo.R./UICI courses in transcription of texts containing formulas;
- Organization of courses for teachers.

Work in progress

- Extension of the package to handle multline environment (align, multline, ...; almost completed);
- Accessible library (L. Pandolfi, Analisi Matematica 1 e 2). We are looking for colleagues that wish to contribute;
- Applications of the Mumford-Shah functional (variational approach);
- Application of time-frequency analysis (Fourier analysis approach);
- Natural Language Generation;
- Human Computer Interaction;
- A virtual Physics Laboratory;
- An accessible Chemistry Laboratory;
- Accessibility of statistical softwares for data analyses.

Our recent papers

1. G. Airò Farulla, T. Armano, A.Capietto, N. Murru, R. Rossini, *Artificial neural networks and fuzzy logic for recognizing alphabet characters and mathematical symbols*, Lecture Notes in Computer Science, Vol. 9759, Proc. "14th International Conference on Computers Helping People with Special Needs", 7-14, 2016.
2. T. Armano, M. Borsero, A. Capietto, N. Murru, A. Panzarea, A. Ruighi,*On the accessibility of Moodle 2 by visually impaired users, with a focus on mathematical content*, Universal Access in the Information Society, DOI 10.1007/s10209-017-0546-8, 2017.
3. T. Armano, A. Capietto, S. C., N. Murru, A. Ruighi, E. Taranto, *An automatized method based on LaTeX for the realization of accessible PDF documents containing formulae*, Lecture Notes in Computer Science, Proc. "16th International Conference on Computers Helping People with Special Needs", 2018.

Package axessibility available at
`https://ctan.org/tex-archive/macros/latex/contrib/axessibility`

Our repository:
`http://www.integr-abile.unito.it/axessibility/?repository`

Thank you very much for your attention!