LaTeX to Braille

Solutions for the automatic transcription of math in braille with 6 and/or 8 points & automatic conversion system from, amongst and for latex to mathml and braille

JULY 2021

PROGETTO DDMATH
Digital learning in mathematics for blind students
ERASMUS+ Program

Co-funded by the Erasmus+ Programme of the European Union

DDMATH

ERASMUS+ Program

DDMATH PROJECT

Digital learning in mathematics for blind students

LaTeX to Braille

Solutions for the automatic transcription of math in braille with 6 and/or 8 points & automatic conversion system from, amongst and for latex to mathml and braille.

July 2021

Project within the ERASMUS+ program – School Education Gateway – Action KA2 - Strategic partnerships for digital education KA226 – Convention n. 2020-1-IT02-KA226-SCH-09557

Project number: ERASMUS+ program – School Education Gateway – Action KA2 - Strategic partnerships for digital education KA226 – Convention n. 2020-1-IT02-KA226-SCH-09557
Project title: DDMATH- Digital learning in mathematics for blind students
IO type: Publication
IO number 1.2
Contractual date of the Delivery: July 2021
Actual date of the Delivery: July 2021
Title of IO: LaTeX to Braille
Nature of the Deliverable: Public
Authors: Associazione PAccini
Abstract: Solutions for the automatic transcription of math in braille with 6 and/or 8 points & automatic conversion system from, amongst and for latex to mathml and braille
Keyword List: Blind, mathematic, Braille, 8dots, education, computing, LaTeX
The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

CONTENTS

1 INTRODUCTION

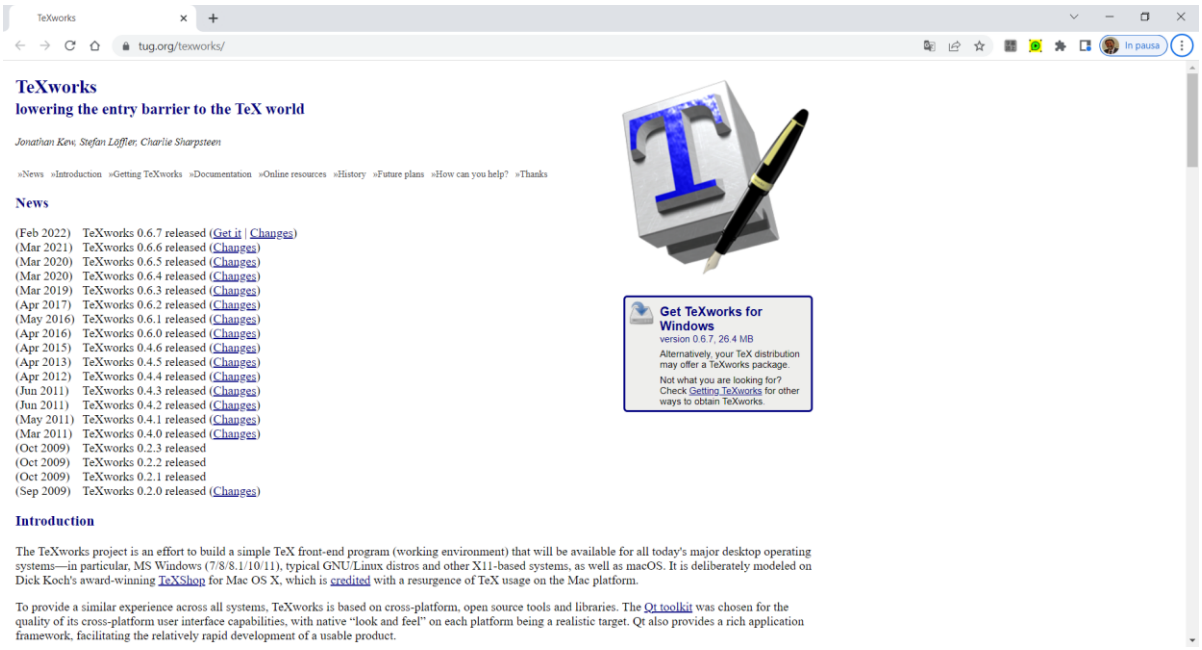
Tex (Knuth, 1987) was developed as a typesetting (markup) code, capable of representing a wide range of printed notation, including mathematics. Let's say right away that it was not born as a notation for blind users. Later, Latex (Lamport, 1988) was developed, based on Tex, and designed to be more easily usable. (Its name derives from 'Tex for Laymen'). LaTeX has been widely adopted by those who write technical texts. Latex is very much geared towards the needs of typographers and not mathematicians, as it was originally intended not as a notation to be read in its raw form by human readers, but to be interpreted by composition software before being read by people.

However, LaTeX has been used extensively by blind mathematicians. There are a number of reasons for this. Firstly, it is very clear, most of the printed mathematical notation can be expressed in Latex. Secondly, it is used extensively by sighted people as a means of carrying out their work. It can therefore be used equally by blind and sighted mathematicians. The sighted can see it as rendered by a typewriter, while the blind can read the raw markup - usually using a computer with a screen reader that renders text to speech or Braille.

Latex is a notation that uses prefixes. It is quite verbose in the sense that it uses words to represent symbols

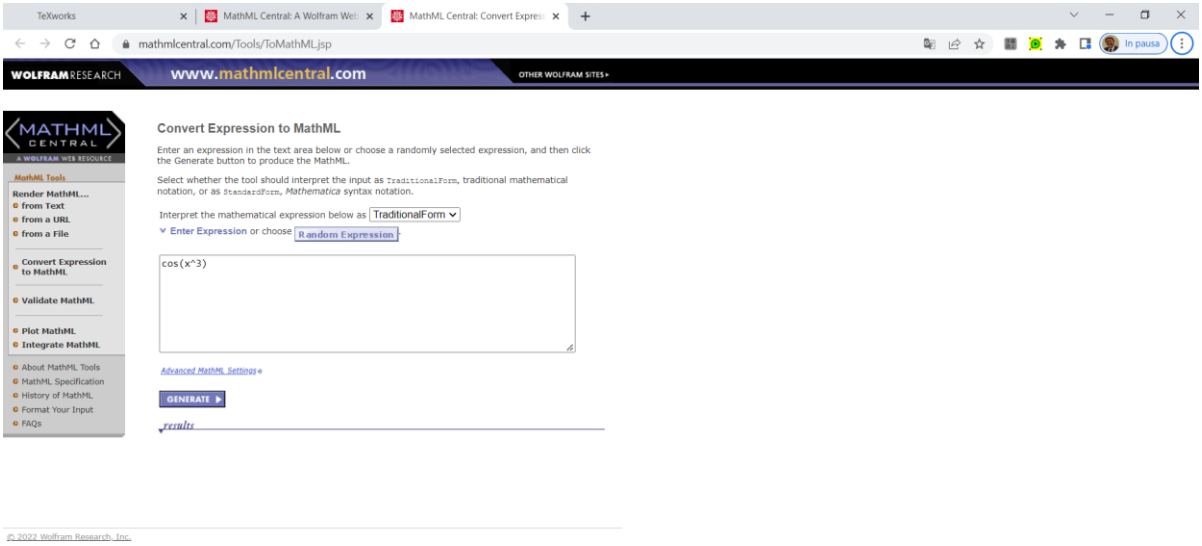
1.1 LATEX https://tug.org/texworks/

Tex is a complex program that resembles a programming language, but actually isn't. What it does is processing a series of strings from an input source and, hopefully, developing it into a well layed-out document to print or look at, on screen. The TeXworks project is an effort to build a simple TeX front-end program (working environment) that will be available for all today's major desktop operating systems. TeXworks includes an integrated PDF viewer.



1.2 MATHML

https://www.mathmlcentral.com/Tools/ToMathML.jsp

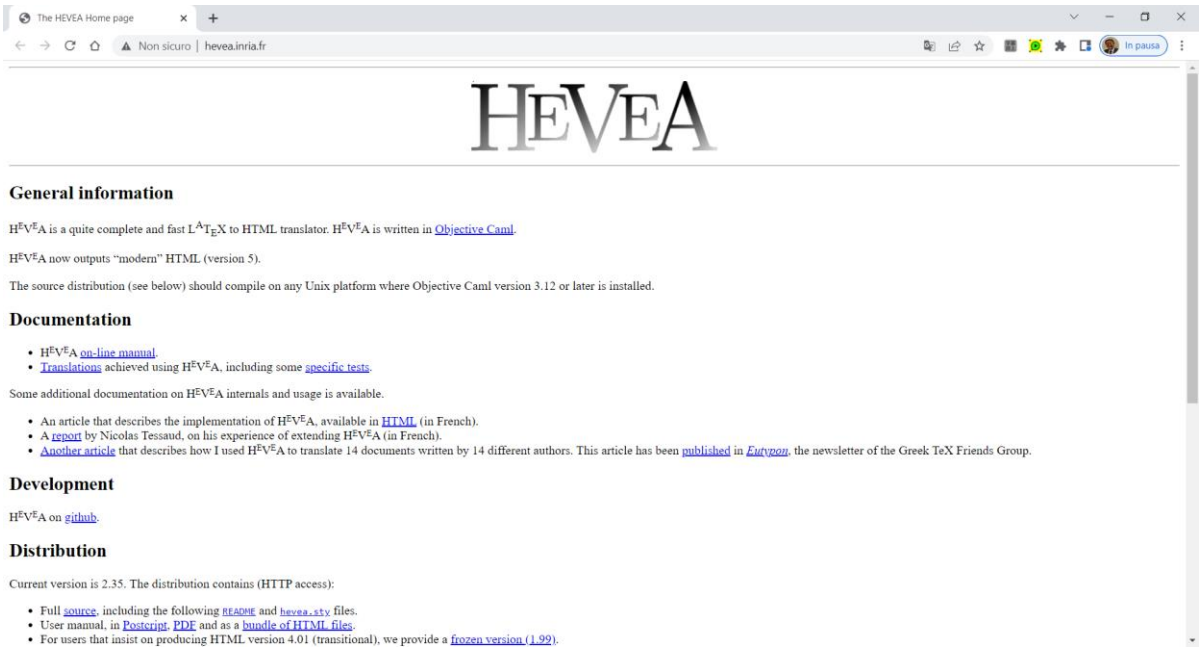


The MathML (acronym of Mathematical Markup Language) is a web language used to represent symbols and math formulas, which also allows to assign a semantic meaning to formulas. It has been derivate from XML, as a specification of the W3C workgroup regarding mathematic. MathML not only deals with the presentation but also with the components meaning of formulas. Numerous programs are available, that can convert math expressions into MathML, that includes also converters between TEX and MathML. Additionally, Wolfram Research make a program that can convert mathematical expressions into MathML. Among the major browsers, those that directly support the format are the recent versions of Mozilla and its derivates, Opera versions starting from 11.60 and Google Chrome, from version 24.

There are external plugins that allow to use the format with also other browsers; Internet Explorer, for example, supports it through MathPlayer. MathML is also supported by office software like Apple suite (Pages, Keynote, Numbers), LibreOffice and Microsoft Word, and by math software, such as Mathematica.

2 RELATED WORK

2.1 HEVEA http://hevea.inria.fr/

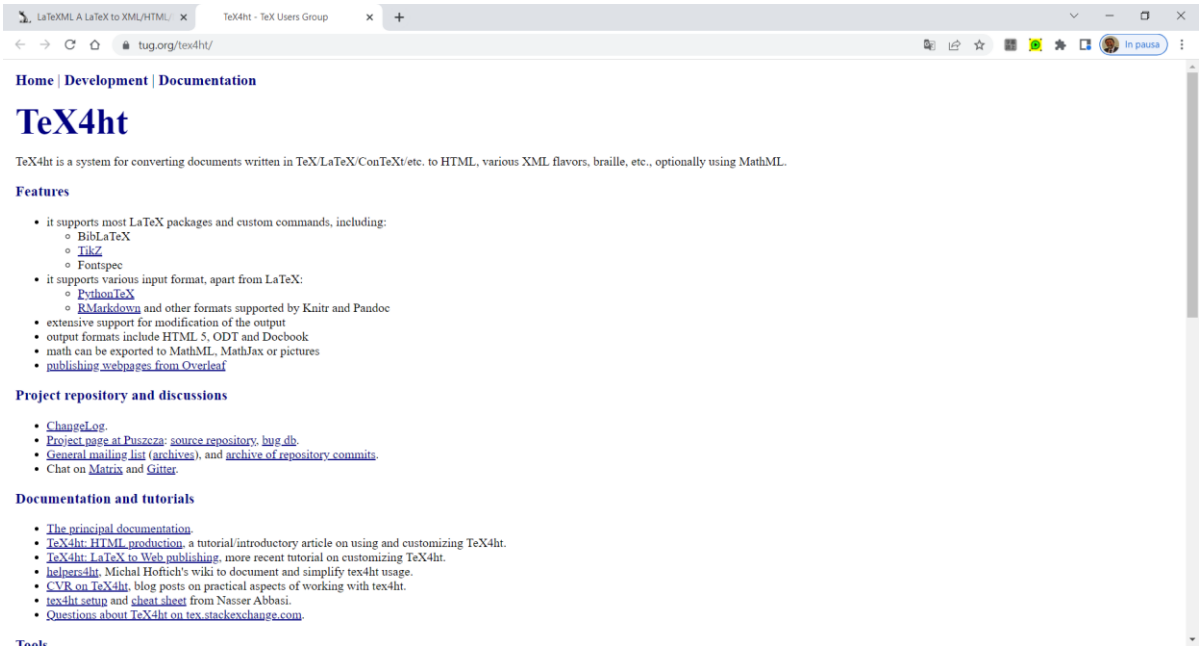


HEVEA is a quite complete and fast LATEX to HTML translator. HEVEA is written in Objective Caml. HEVEA now outputs “modern” HTML (version 5). The source distribution should compile on any Unix platform where Objective Caml version 3.12 or later is installed.

2.2 LATEXML https://math.nist.gov/~BMiller/LaTeXML/

The approach is to emulate TeX as far as possible (in Perl), converting the TeX or (LaTeX) document into LaTeXML's XML format. That format is modelled on the typical document structure found in LaTeX, and inspired by HTML, MathML, OpenMath and others. That abstract document is then further transformed into HTML of various flavors, with MathML and SVG, or into JATS or ePub or …. Of course, emulating TeX is kinda hard, there are many clever LaTeX package developers, and the Web moves quickly, so there are gaps in fidelity and coverage.

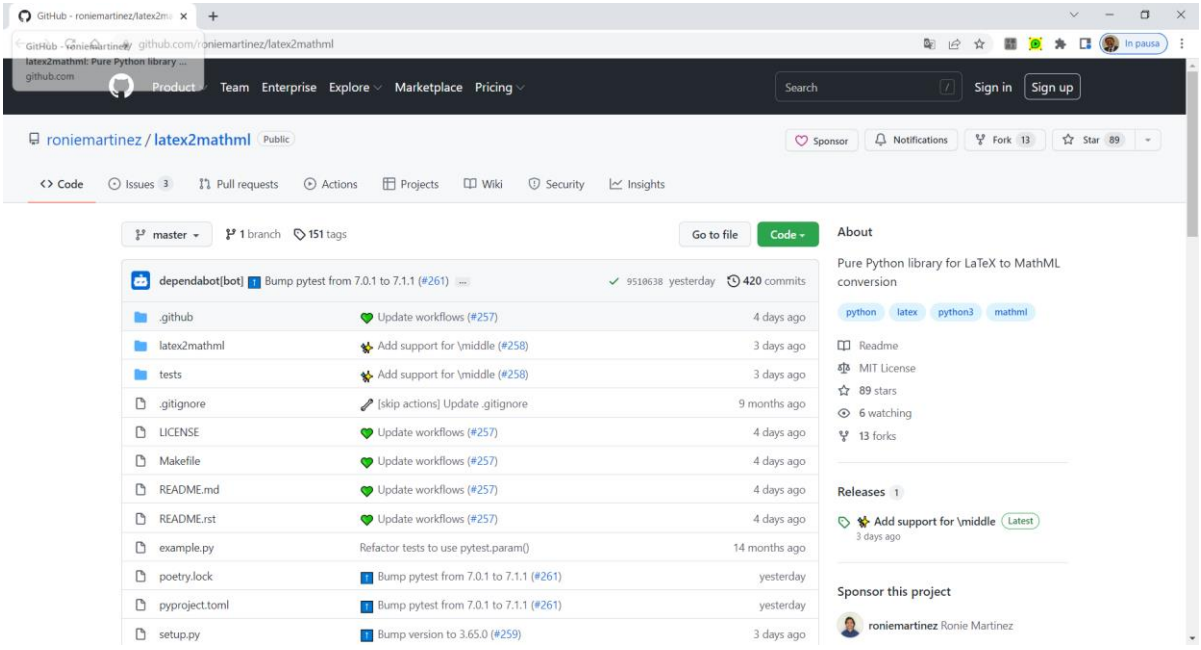
2.3 TEX4HT https://www.tug.org/tex4ht/



TeX4ht is a system for converting documents written in TeX/LaTeX/ConTeXt/etc. to HTML, various XML flavors, braille, etc., optionally using MathML.

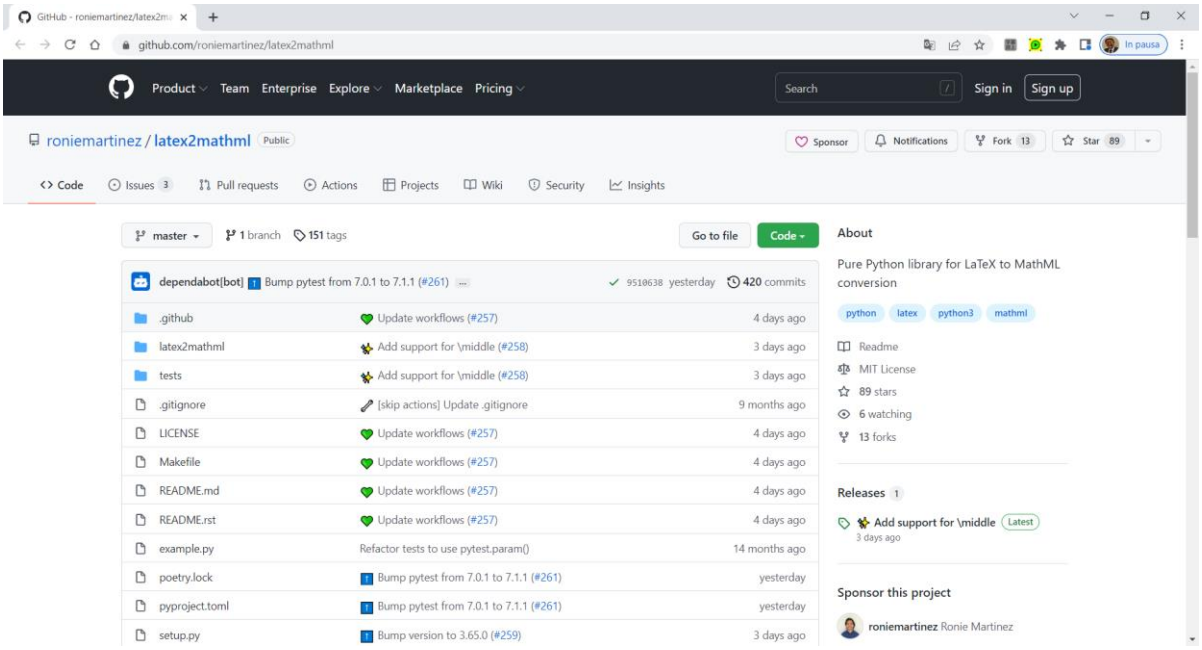
2.4 LATEX2MATHML

https://github.com/roniemartinez/latex2mathml



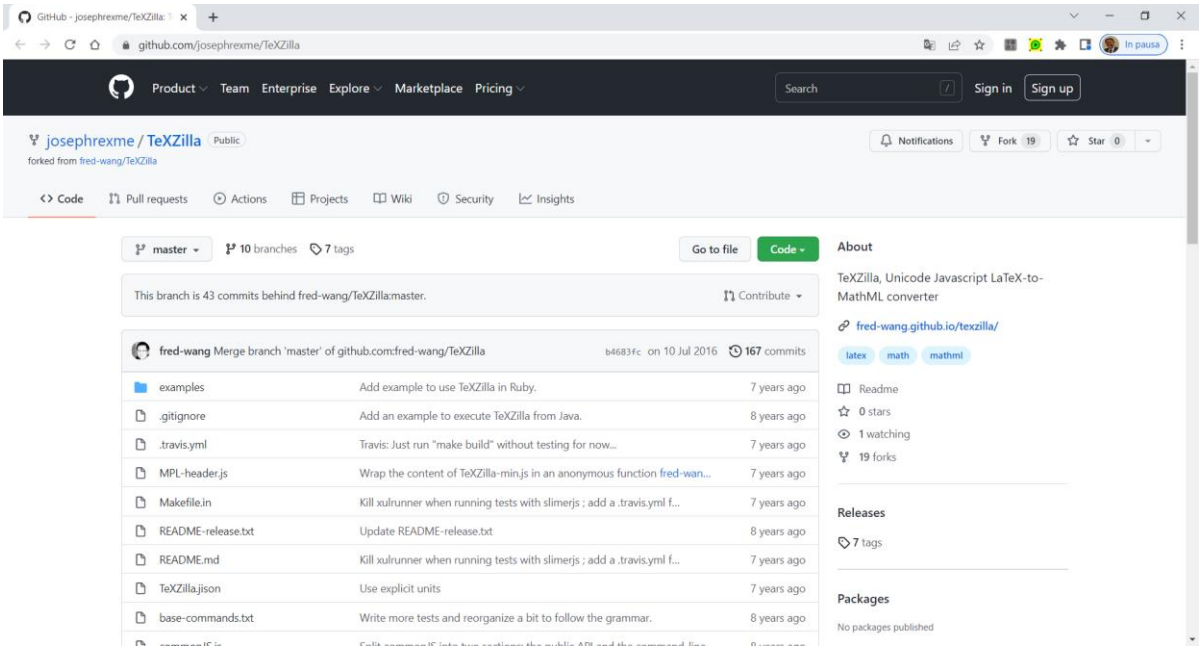
Pure Python library for LaTeX to MathML conversion.

2.5 MATHCONVERTER https://github.com/oerpub/mathconverter



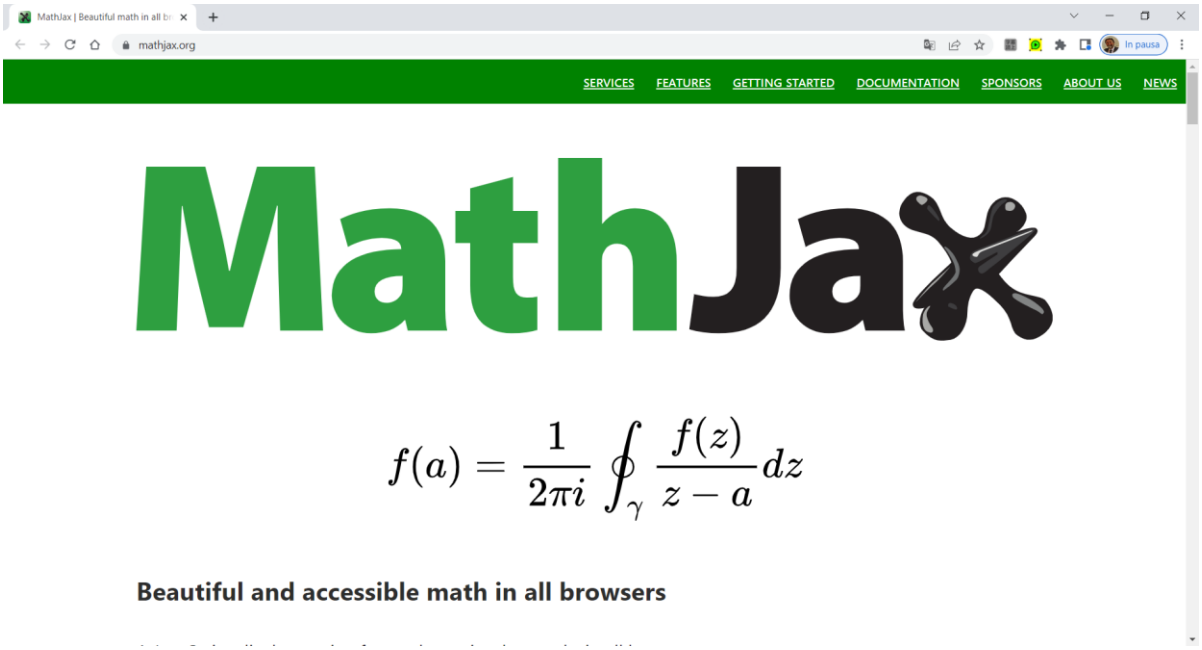
Converts from AsciiMath, LaTeX, MathML to LaTeX, MathML utilizes MathMLCloud (for MathML output) and XSL transforms.

2.6 TEXZILLA https://github.com/josephrexme/TeXZilla



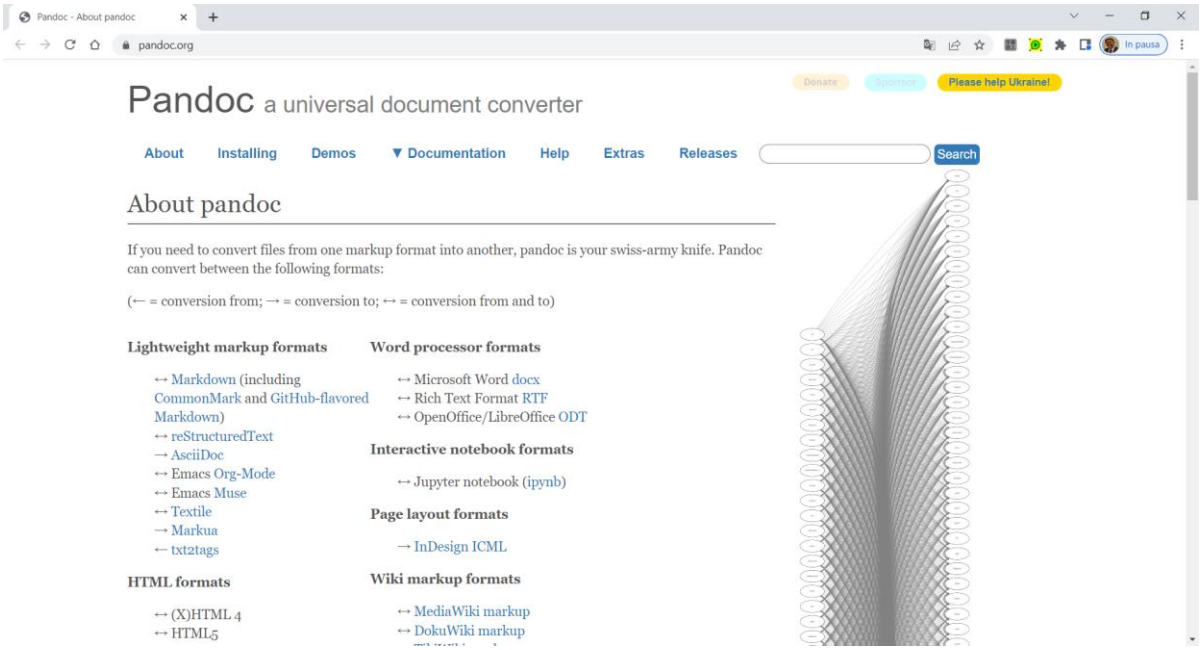
TeXZilla is a Javascript LaTeX-to-MathML converter compatible with Unicode. This is still a work in progress and things may change in the future.

2.7 MATHJAX https://www.mathjax.org/



A JavaScript display engine for mathematics that works in all browsers. No more setup for readers. MathJax is highly modular on input and output. Use MathML, TeX, and ASCIImath as input and produce HTML+CSS, SVG, or MathML as output.

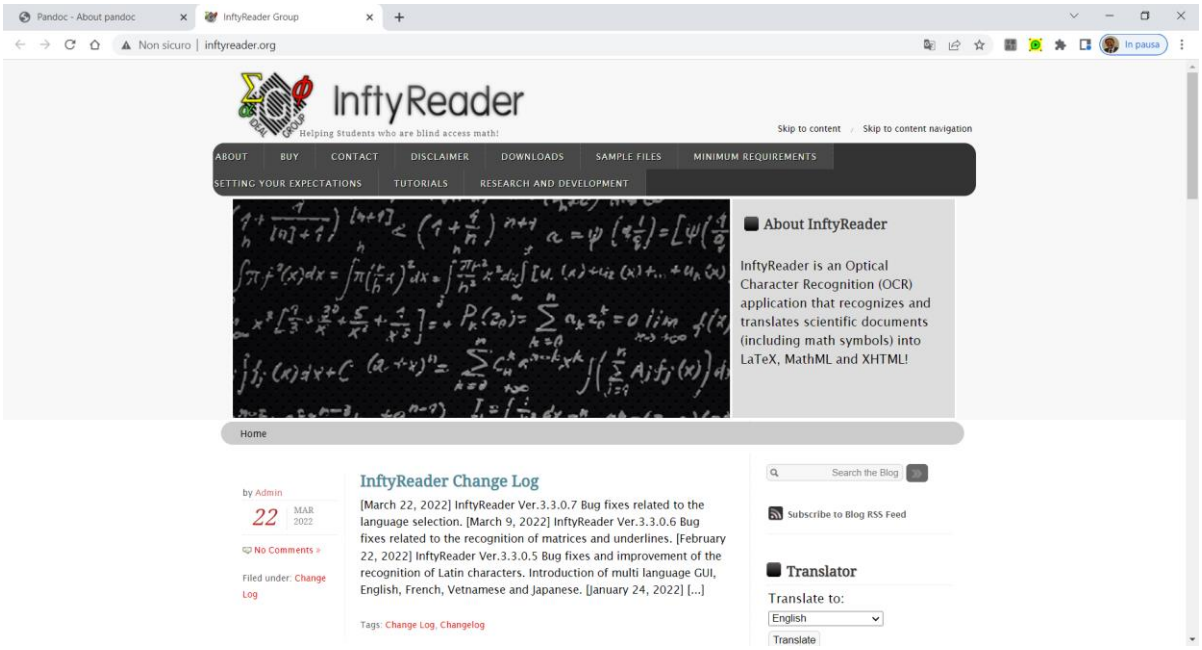
2.8 PANDOC https://pandoc.org/



Pandoc understands a number of useful markdown syntax extensions, including document metadata (title, author, date); footnotes; tables; definition lists; superscript and subscript; strikeout; enhanced ordered lists (start number and numbering style are significant); running example lists; delimited code blocks with syntax highlighting; smart quotes, dashes, and ellipses; markdown inside HTML blocks; and inline LaTeX. If strict markdown compatibility is desired, all of these extensions can be turned off. LaTeX math (and even macros) can be used in markdown documents. Several different methods of rendering math in HTML are provided, including MathJax and translation to MathML. LaTeX math is converted (as needed by the output format) to unicode, native Word equation objects, MathML, or roff eqn.

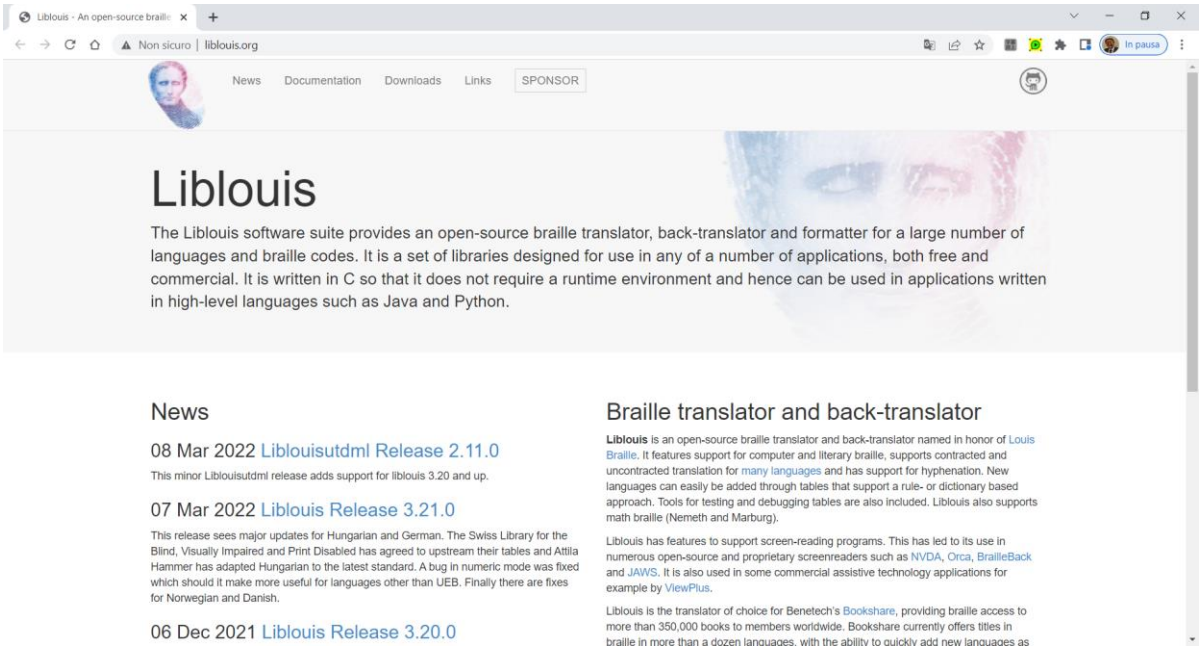
2.9 INFTYREADER http://www.inftyreader.org/

https://www.sciaccess.net/en/InftyReader/



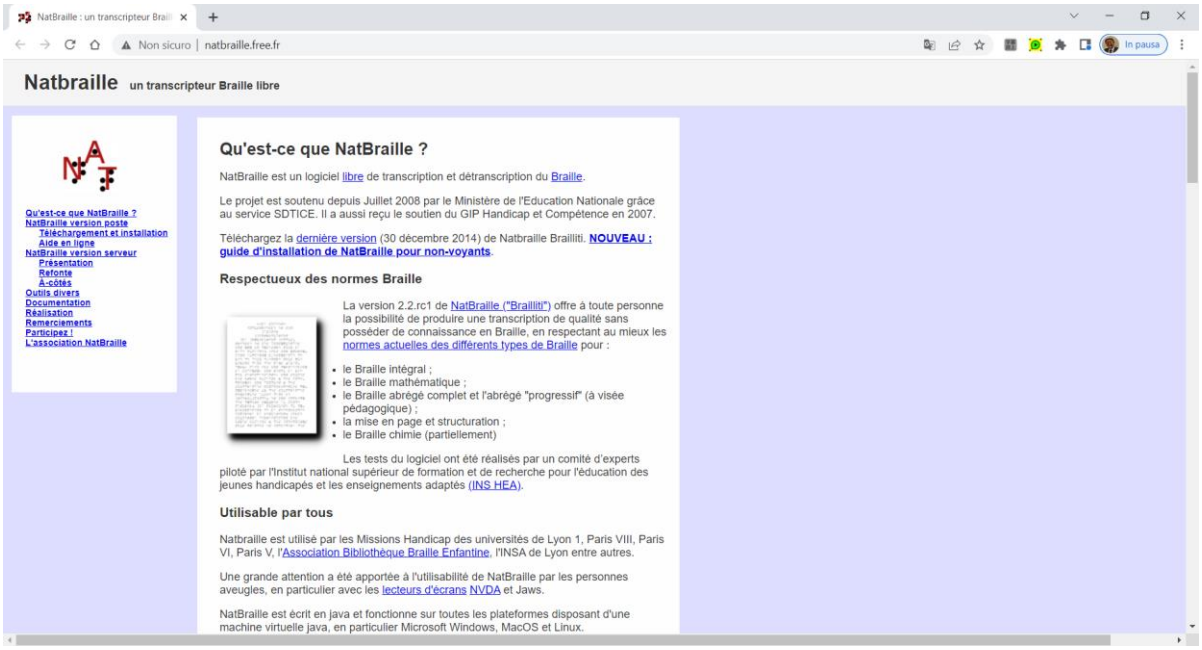
InftyReader is OCR software to recognize scientific documents including mathematical formulae. "InftyReader" converts PDF and scanned images to various types of accessible documents: LaTeX, XHTML(MathML), HRTeX, IML, Microsoft Word document, EPUB3, PDF with TeX and Chattybook (Audio HTML). For the scanned image files or Image PDF produced from scanned images, InftyReader uses OCR specially trained for STEM documents recognizing special math symbols and analyzing math structures. For e-born PDF, InftyReader uses a PDF parser rather than OCR, so the character recognition results are very accurate, not only for ordinary texts but also math symbols.

2.10 LIBLOUIS http://liblouis.org/



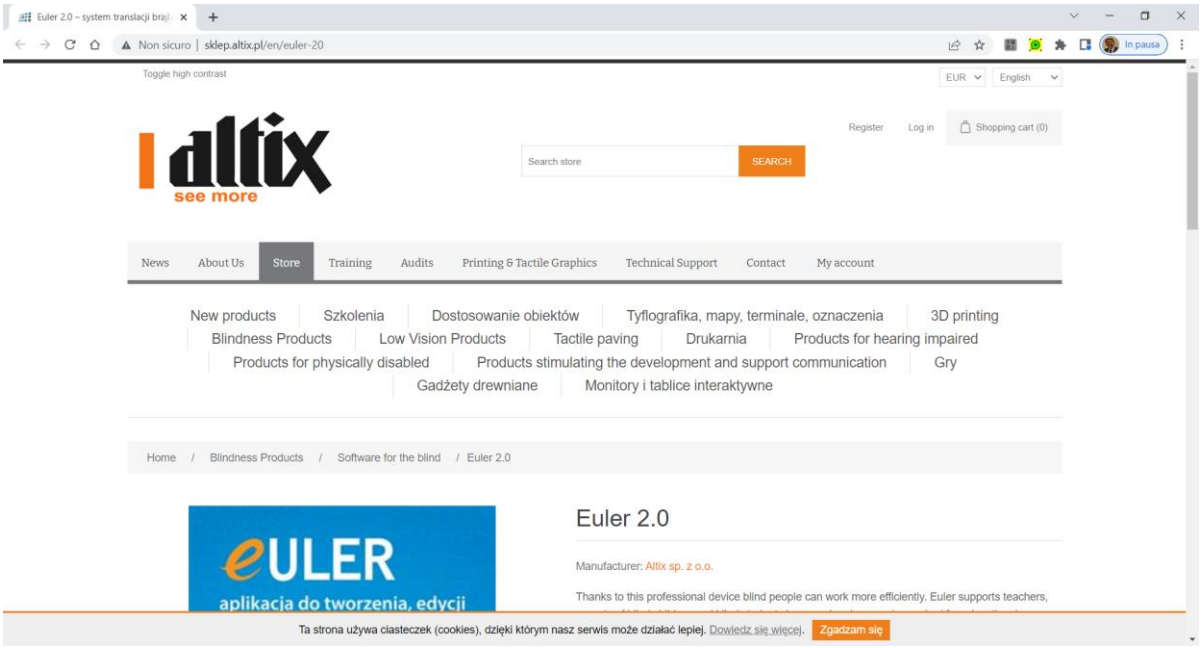
The Liblouis software suite provides an open-source braille translator, back-translator and formatter for a large number of languages and braille codes. It is a set of libraries designed for use in any of a number of applications, both free and commercial. It is written in C so that it does not require a runtime environment and hence can be used in applications written in high-level languages such as Java and Python.

2.11 NATBRAILLE http://natbraille.free.fr/



NatBraille is free Braille transcription and detranscription software. The project has been supported since July 2008 by the Ministry of National Education thanks to the SDTICE service. It also received support from GIP Handicap et Competence in 2007. NatBraille can transcribe or detranscribe the following input formats: Documents in OpenDocument format such as those produced by LibreOffice or OpenOffice, Documents produced by Microsoft Word, Documents in text format that may contain mathematical expressions written in the MathML language Simple, standards-compliant HTML documents. Braille documents in text format for transcription.

2.12 EULER 2.0 http://sklep.altix.pl/en/euler-20



Euler 2.0 is a professional converter offering advanced editing options both in braille and regular printing formats. Thanks to the Euler functionalities users can write texts using Braille keyboards, and Euler translates Braille format into regular text, or one can translate regular text into braille format. The mathematical documents can be printed in each of these possible formats: Epheser, Nemeth Code and others. Euler enables users to read, edit and create documents in many formats such as: DOC, DOCX, RTF, HTML, PDF, XLS, XLSX, and TXT. The braille document can be saved as RTF, BRL or even XLS.

3 EXAMPLE

In these image sequences we present the steps of how to transcribe a page from Latex into Lambda Braille.

The first image shows a page written in LateX.

File TeX:

```
\documentclass[a4paper]{article}
\usepackage[italian]{babel}
\usepackage[utf8]{inputenc}
\usepackage[colorinlistoftodos]{todonotes}
\usepackage{amsthm}
\usepackage{amsmath}
\usepackage{siunitx} % gradi

\title{Esercizi}
\author{Liceo Scientifico}
\date{}

\begin{document}
\maketitle
\hrule

\section{Trigonometria}

Risolvere i seguenti esercizi:
\begin{itemize}
    \item $\sin (2x) = \sqrt  2  \cos x$
    \item $1  + 2 \cos^2 x = 2\sqrt 3 \sin x \cos x$
    \item $\cos (3x) - \sin (3x) < 1$
    \item $\dfrac{\sin x}{2 \sin^2 x  - 3 \cos x} \geq 0$
\end{itemize}

\bigskip
\noindent Risolvere il seguente problema:
\bigskip

\noindent Si vuole misurare la distanza tra due alberi, rappresentati in figura dai due punti $C$ e $D$. Non è possibile però farlo direttamente perchè i due alberi sono separati dai due osservatori (rappresentati dai punti $A$ e $B$) da un fiume invalicabile. Il primo osservatore, posto in $A$, misura i due angoli $B\hat AC$ e $C\hat AD$ : trova che essi misurano rispettivamente $\ang{40}$ e $\ang{24}$. Il secondo osservatore, posto in $B$, alla distanza di $100 m$ da $A$, misura i due angoli $A\hat BD$ e $C\hat BD$: trova che essi misurano rispettivamente $\ang{52}$ e $\ang{80}$. Calcola:
\begin{itemize}
    \item la distanza tra $A$ e $D$.
    \item la distanza tra i due alberi (ossia la distanza tra $C$ e $D$).
\end{itemize}
```

The page as it appears in PDF format.

File PDF:

Esercizi

Liceo Scientifico

1 Trigonometria

Risolvere i seguenti esercizi:

- $\sin(2x) = \sqrt{2}\cos x$
- $1 + 2\cos^2 x = 2\sqrt{3}\sin x \cos x$
- $\cos(3x) - \sin(3x) < 1$
- $\dfrac{\sin x}{2\sin^2 x - 3\cos x} \geq 0$

Risolvere il seguente problema:

Si vuole misurare la distanza tra due alberi, rappresentati in figura dai due punti C e D. Non è possibile però farlo direttamente perchè i due alberi sono separati dai due osservatori (rappresentati dai punti A e B) da un fiume invalicabile. Il primo osservatore, posto in A, misura i due angoli $B\hat{A}C$ e $C\hat{A}D$: trova che essi misurano rispettivamente $40°$ e $24°$. Il secondo osservatore, posto in B, alla distanza di $100m$ da A, misura i due angoli $A\hat{B}D$ e $C\hat{B}D$: trova che essi misurano rispettivamente $52°$ e $80°$. Calcola:

- la distanza tra A e D.
- la distanza tra i due alberi (ossia la distanza tra C e D).

The image shows the LaTeX to MathML conversion, using Pandoc TeX to MathML.

Pandoc TeX to MathML:

```
<h1 class="title">Esercizi</h1>
<p class="author">Liceo Scientifico</p>
</header>
<hr />
<h1 id="trigonometria">Trigonometria</h1>
<p>Risolvere i seguenti esercizi:</p>
<ul>
<li>
  <p><math display="inline" xmlns="http://www.w3.org/1998/Math/MathML">
        <semantics>
            <mrow>
              <mo>sin</mo>
              <mrow>
                <mo stretchy="true" form="prefix">(</mo><mn>2</mn><mi>x</mi><mo stretchy="true" form="postfix">)</mo>
              </mrow>
              <mo>=</mo>
              <msqrt><mn>2</mn></msqrt>
              <mo>cos</mo><mi>x</mi>
            </mrow>
            <annotation encoding="application/x-tex">\sin (2x) = \sqrt  2  \cos x</annotation>
        </semantics>
      </math>
    </p>
</li>
```

From the MathML format the page is imported in 8-dot Lambda Braille.

File Lambda: