

latex2nemeth: A direct L^AT_EX-to-Braille Transcribing Tool

Andreas Papasalouros and Antonis Tsolomitis

Department of Mathematics
University of the Aegean

TUG 2021

Contents

- Motivation, aims of `latex2nemeth`
- The Braille and Nemeth codes
- Characteristics of the `latex2nemeth` tool
- Implementation
- Evaluation and conclusion

Motivation

- In 2014 we came to know about a visually impaired student of Mathematics at the University of Athens.
- The student could read and write in Braille notation and in the Nemeth representation of Mathematics.
- Although the lecture notes and course textbooks were available in \LaTeX source format, there was no reliable way to translate these sources into tactile representations, accessible by visually impaired persons.
- The student had to manually transcribe learning content with the help of a seeing person.
- We set out to create a tool for transcribing \LaTeX texts to tactile representations accessible to blind persons.

Motivation

- In 2014 we came to know about a visually impaired student of Mathematics at the University of Athens.
- The student could read and write in Braille notation and in the Nemeth representation of Mathematics.
- Although the lecture notes and course textbooks were available in \LaTeX source format, there was no reliable way to translate these sources into tactile representations, accessible by visually impaired persons.
- The student had to manually transcribe learning content with the help of a seeing person.
- We set out to create a tool for transcribing \LaTeX texts to tactile representations accessible to blind persons.
- The above is an instance of a wider problem: Visually impaired students of related fields did not have access to the bulk of study material available in \LaTeX format.

The Braille code

- The Braille system allows the tactile representation of characters in various alphabets, giving access to reading texts to visually impaired persons.
- The six-dot Braille system supports the representation of $2^6 = 64$ different characters.
- An assignment of Braille symbols to letters (encoding) defines a certain Braille alphabet (eg, English, Greek, etc.).

Reading of Braille/Nemeth code



A Braille embosser



The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols,

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets,

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets, fractions,

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets, fractions, superscripts and subscripts,

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets, fractions, superscripts and subscripts, radicals,

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets, fractions, superscripts and subscripts, radicals, symbols of grouping,

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets, fractions, superscripts and subscripts, radicals, symbols of grouping, spatial arrangements,

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets, fractions, superscripts and subscripts, radicals, symbols of grouping, spatial arrangements, etc.

The Nemeth code

- The Nemeth code for Mathematics allows the representation of Mathematics symbols and expressions into the Braille system.
- Defines a set of *rules* that designate the combinations of Braille symbols that describe various types of mathematical structures.
- The original document of Nemeth (1972) specifies 25 rules covering various aspects of mathematical structures such as:
numeric signs and symbols, alphabets, fractions, superscripts and subscripts, radicals, symbols of grouping, spatial arrangements, etc.
- Available at
https://nfb.org/images/nfb/documents/pdf/nemeth_1972.pdf

Aims of the `latex2nemeth` project

- The *reliable* transcription of books and electronic notes from \LaTeX to Nemeth/Braille.
- The creation of a repository of mathematical texts available to visually impaired students and researchers.
- At least at the beginning of this project (end of 2014), to the best of our knowledge, no tool or method met the above requirements.

Existing solutions

- Commercial solutions
 - Conversion from \LaTeX to Word with MathType.
 - Conversion from Word to Braille with Duxbury (with limitations).
 - Official support only for a small subset of \LaTeX commands.
 - The above process produced unreadable Braille code with many mistakes (extensive tests during 2014).
- Open source solutions (liblouis)
 - Conversion from \TeX into MathML (e.g. with tex4ht).
 - Then, conversion from MathML into Braille with the liblouis library.
 - The process creates Braille code with many errors.
 - The liblouis library did not aim at supporting \TeX at the time of the creation of the `latex2nemeth` program (extensive tests during 2014).

Program features

- \LaTeX files with text in Greek, English and Ancient Greek are converted to Braille.
- More than 850 different mathematical symbols and expressions are supported.
- All \TeX AMS mathematical symbols are covered, among others.
- Different Braille alphabets can be supported with the use of different *symbol tables*.
 - A symbol table is a JSON file that maps individual characters or \TeX commands to Braille characters or sequences of characters.

Flow of translation to Braille/Nemeth

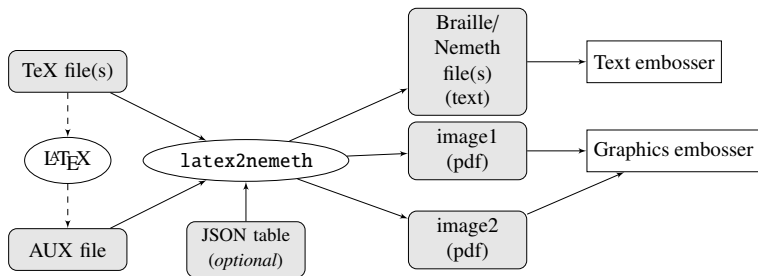


Image label filtering

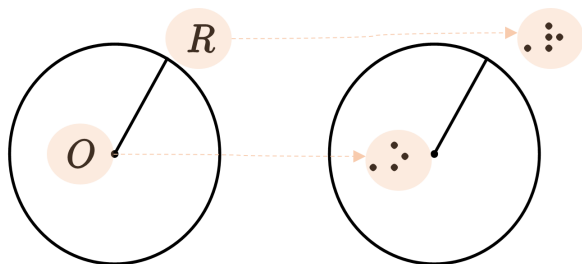
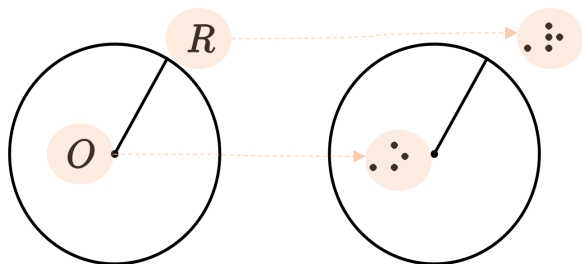


Image label filtering



Original code

```
\begin{pspicture}(-1,-1)(1.2,1.2)
\pscircle(0,0){1}
\qdisk(0,0){1pt}
\uput[180](0,0){$O$}
\psline(0,0)(.5,.866)
\uput[30](.5,.866){$R$}
\end{pspicture}
```

Transcribed code

```
\begin{pspicture}(-1,-1)(1.2,1.2)
\pscircle(0,0){1}
\qdisk(0,0){1pt}
\uput[180](0,0){.:.}
\psline(0,0)(.5,.866)
\uput[30](.5,.866){.:.}
\end{pspicture}
```

Implementation

- The transcriber is based on a parser for the \LaTeX language, created from scratch.
- The parser recognizes most of the most common \LaTeX commands and environments in text and mathematics modes, supporting both English and Greek characters and also covers most structures and mathematical symbols.
- The program is developed in Java.
- The JavaCC compiler generation tool for the generation of the lexical analyzer and the parser.
- The tool is available as a jar (Java archive) package.

Implementation: L^AT_EX to Braille transcription process

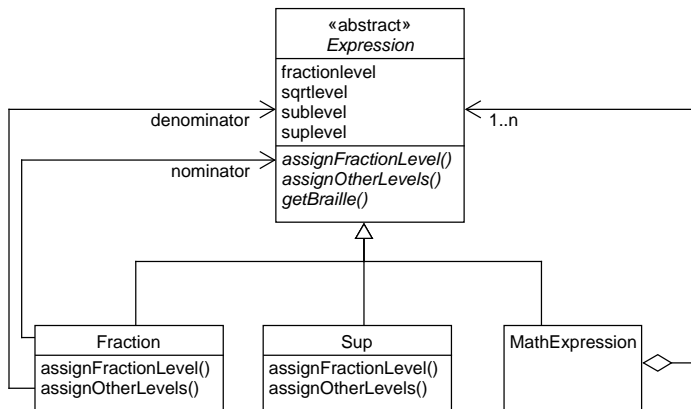
- Each paragraph and each environment in the input L^AT_EX sources is processed separately.
- In text mode, each lexical token is recognized and transcribed into its corresponding Braille symbol by using a certain *symbol table*.
- Numerical expressions are lexically scanned as atomic entities , eg. number 13.455, since, according to Braille code, a certain number indicator (.:) must precede the whole numerical expression.

Parsing of mathematical expressions

- Mathematical expressions are parsed into appropriate syntactical trees in memory.
- The abstract syntax trees for mathematics expressions are independent of the target language (Nemeth).
- Depending on the type of expression (fraction, superscript/subscript, etc) an appropriate procedure (semantic routine) generates corresponding Nemeth code.

Implementation

- An object-oriented representation of mathematical expressions was adopted, based on the Composite and Interpreter design patterns (Gamma et al., 1996).



A sample expression

$$e^{x^{b+1}} + \frac{c}{d + \frac{k}{x+3}}$$

A sample expression

The above expression is rendered in Nemeth Braille code as

$$e^{x^{b+1}} + \frac{c}{d + \frac{k}{x+3}}$$

A sample expression

The above expression is rendered in Nemeth Braille code as

$$e^{x^{b+1}} + \frac{c}{d + \frac{k}{x+3}}$$

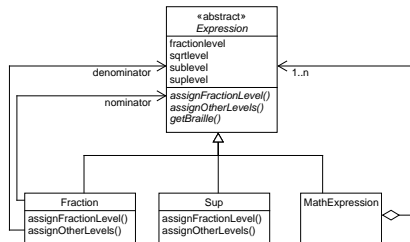
```
e<exp>x<exp-2>b+1<base>  
+  
<openfrac-2>  
c  
<fractionbar>  
d  
+  
<openfrac>k<fractionbar>x+3<closefrac>  
<closefrac>
```

Expression indicators and depth

- In the example, the fraction and the exponent are defined by appropriate *expression indicators*.
- In order to assist a blind person to conceive the structure of the expression, an expression indicator signifies the *depth* of the expression. Thus,
 - `<exp>` (`:`) signifies a simple superscript
 - `<exp-2>` (`:` `:`) signifies a superscript within a superscript.
 - `<openfrac>` (`::`) signifies the opening of a single fraction.
 - `<openfrac-2>` (`.` `::`) signifies the opening of a complex fraction, etc.

Expression generation

- Depth indicators for various expressions is calculated by appropriate methods of the instances of the Expression class.
- Depth for complex fractions is calculated *bottom-up* by the `assignFractionLevel()` method.
- Depth for other nested expressions (superscripts, subscripts, roots, etc) is calculated *top-down* by the `assignOtherLevels()` method.



Spatial arrangements

Support for spatially aligned structures such as arrays, cases, align expressions, etc.

Spatial arrangements

Support for spatially aligned structures such as arrays, cases, align expressions, etc.

```
$$\chi_F (x)=\left\{\begin{array}{ll} 1, & \text{\textit{if } } x\in F\\ 0, & \text{\textit{if } } x\notin F \end{array}\right\}.
```

Spatial arrangements

Support for spatially aligned structures such as arrays, cases, align expressions, etc.

```
$$\chi_F (x)=\left\{\begin{array}{l} 1, & \text{\textit{if } } x\in F\\ 0, & \text{\textit{if } } x\notin F \end{array}\right. $$
```

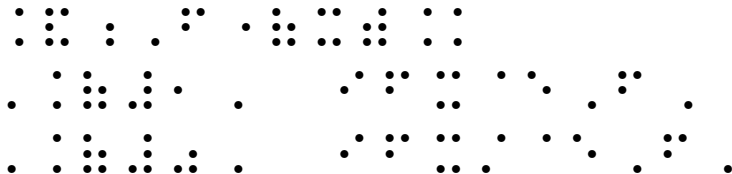
$$\chi_F(x) = \begin{cases} 1, & \text{if } x \in F \\ 0, & \text{if } x \notin F \end{cases}$$

Spatial arrangements

Support for spatially aligned structures such as arrays, cases, align expressions, etc.

```
$$\chi_F(x)=\left\{\begin{array}{l} 1, & \text{\textit{if } } x\in F \\ 0, & \text{\textit{if } } x\notin F \end{array}\right. $$
```

$$\chi_F(x) = \begin{cases} 1, & \text{if } x \in F \\ 0, & \text{if } x \notin F \end{cases}$$



Extending the Nemeth code - proposing new symbols

- Nemeth code prescribes certain mechanisms for defining new symbols.
- As an example, \hookrightarrow (`\hookrightarrow`) is defined based on the plain arrow, expressed in Nemeth as $\cdot\cdot\cdot\cdot$:
 - Character $\cdot\cdot$ signifies grouping of characters in a new symbol.
 - The combination $\cdot\cdot\cdot$ signifies the characteristic hook of the above symbol:
 $\cdot\cdot\cdot\cdot\cdot\cdot\cdot$
- Other new Braille mathematical symbols have been proposed.

Composed Braille/Nemeth symbols

\LaTeX command	symbol	composition	braille symbol
<code>\Cap</code>	\cap	\cap inside \cap	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\circledcirc</code>	\odot	\circ inside circle	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\leadsto</code>	\curvearrowright	wavy arrow shaft+arrow	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\multimap</code>	\multimap	arrow soft ending in \circ	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\succsim</code>	\succsim	$>$ with \sim below	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\subsetneqq</code>	\subsetneqq	\subset but \neq	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\lesseqgtr</code>	\lesseqgtr	less or equal or greater	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\gtrsim</code>	\gtrsim	approximately greater	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\ggg</code>	\ggg	much greater	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\triangleq</code>	\triangleq	equal with triangle	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\lessdot</code>	\lessdot	less with dot inside	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\eqslantless</code>	\eqslantless	less with equal above	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\Vvdash</code>	\Vvdash	triple vertical line dash	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\approxq</code>	\approxq	approximately equal	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\nsim</code>	\nsim	not equivalent	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\varsubsetneqq</code>	\varsubsetneqq	subset but not equal	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\varsubsetneq</code>	\varsubsetneq	subset but not equal	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\precnsim</code>	\precnsim	variant of less but not equal	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$
<code>\succapprox</code>	\succapprox	variant of greater but not equal	$\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot$

Documentation

- Available at the web site
 - Dictionary of mathematical terms (from symbol name to Nemeth).
 - Reverse dictionary of mathematical terms (from Nemeth to symbol name).
 - The above dictionaries are available for both seeing and visually impaired persons.
- Scientific papers
 - Andreas Papasalouros, Antonis Tsolomitis, A Direct TeX-to-Braille Transcribing Method, *Journal of Science Education for Students with Disabilities*, vol. 20, no. 1, pp. 36–49, RIT Scholar works, 2017.
 - Andreas Papasalouros, Antonis Tsolomitis, A Direct TeX-to-Braille Transcribing Method. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pp. 373–374, ACM Press, 2015.

Distribution

- Web page of the project
`https://myria.math.aegean.gr/braille/index-en.html`
- CTAN repository
`https://ctan.org/pkg/latex2nemeth`
- Source code
`https://sourceforge.net/projects/latex2nemeth/`
- `latex2nemeth` is currently available in TeX Live, MiKTeX and MacTeX distributions.

Towards a repository for mathematical/scientific content for the visually impaired

- We are creating a repository of publicly available mathematical/scientific content in Braille code.
- So far, the following complete texts have been transcribed:
 - 1 book in English
 - 12 books and course notes in Greek
 - 1 book in ancient Greek: Homer's Odessey
- All books are available at the web page of the project.

Evaluation – Reliability of transcription

- A formal evaluation has been conducted by using mathematical documents from AMS.
- The quality of the transcription was evaluated by comparing the transcribed documents with back translations by two visually impaired persons with good knowledge of the Nemeth code.
- The interrater agreement among the back translations of the two participants was very high (Cohen's kappa ≈ 0.98).
- Informal evaluation: Much of the generated content has been used as study material by a blind student (mentioned before).
- The student has graduated on time and now she is pursuing graduate studies with the help of `latex2nemeth`.

Limitations and Future work

- The tool has limited support for macros.
- Currently a simple parameter substitution mechanism is implemented.
- We plan a full re-implementation with full macro expansion support.
- The tool fails to discriminate among symbols *period* and *decimal point* (.).
 - The above issue is to be resolved in a forthcoming version.
- We are working on a more extensive support for images.
 - Currently a simple filtering of image labels for PSTricks images is supported.

Acknowledgements

- We would like to thank Mrs. Olga Maleza for the support and the information that she has provided from the beginning of this project.
- The development of the tool was partially supported by the University of the Aegean Research Unit (grant no. 2625).

Thank you!