

Article

# An Assistive Technology for Braille Users to Support Mathematical Learning: A Semantic Retrieval System

Zahra Asebriy <sup>1,\*</sup>, Said Raghay <sup>1</sup> and Omar Bencharef <sup>2</sup>

<sup>1</sup> Faculty of Sciences and Techniques, Cadi Ayyad University, Marrakesh 40000, Morocco; s.raghay@uca.ma

<sup>2</sup> Higher School of Technology, Cadi Ayyad University, Essaouira 44000, Morocco; bencharef98@gmail.com

\* Correspondence: za.asebriy@uca.ma

Received: 14 September 2018; Accepted: 19 October 2018; Published: 26 October 2018



**Abstract:** Mathematical learning from digital libraries and the web is a challenging problem for people with visual impairments and blindness. In this paper, we focus on developing the mathematical learning skills of braille users with a new assistive technology developed to retrieve semantically mathematical information from the web. This kind of research is still in the study phase. This paper presents an overview of assistive technologies for braille users followed by a description of the proposed system, which works in four phases. In the first phase, we translate a query math formula in braille into MathML code, and then we extract the structural and semantic meaning from the MathML expressions using multilevel presentation. In the classification phase, we choose a multilevel similarity measure based on K-Nearest Neighbors to evaluate the relevance between expressions. Finally, the query result is converted to braille math expressions. Experiments based on our database show that the new system provides more efficient results in responding to user queries.

**Keywords:** braille code; mathematical-expression retrieval; MathML; assistive technologies

## 1. Introduction

Assistive technologies help people with visual impairments to improve their lives. They remove many barriers to use computer technologies for studying, reading, writing digital documents, interacting with each other, and retrieving information on the web and/or in digital libraries. People with visual impairments use speech synthesizers to access the displayed information on the computer screen or braille display (e.g., MAVIS, LaBraDoor, LAMBDA) to write electronic documents, send emails, and take notes, or hearing by screen readers (e.g., Safari+ VoiceOver, Chrome, MathPlayer, MathTalk, JAWS). However, these assistive technologies have some limitations to access all kinds of information like images, tables, graphs, and mathematical notations [1,2].

People with visual impairments face serious difficulties in dealing with and studying mathematics and sciences. They are limited in accessing printed and digital formats of mathematical documents, as math expressions are highly symbolic and nonlinear. Having access to the content of digital mathematical documents (Word, PDF, and web pages) should be in braille or by voice to be easily used by people with sight disabilities in their studies and communications.

Another constraint for people with visual impairments is that mathematical braille code is not universal, as literary braille and each country has its own code. To illustrate, Nemeth [3] and GS Braille are used in North America, India, Malaysia, and other countries. RNIB (Royal National Institute for the Blind) [4] is used in the United Kingdom, Marburg in Germany, and CMU (In English: Unified Mathematical Code) in Spain, while Middle Eastern countries use Nemeth Braille. Braille code is not a language and it has limits in the conversion of certain information like mathematical notations. As opposed to text, mathematical expressions are not directly convertible into braille code because mathematical notations have two dimensional writings, while braille code is one-dimensional.

Similarly, markup language mathematics such as LaTeX [5] and MathML [6] are used to standardize and normalize mathematical expressions in order to facilitate the conversion into braille.

An additional constraint is the complexity of math expressions and the definition of similarity in math [7]. For example:  $a + b$  is identical to  $x + y$ , and  $3x^2 + x + 1$  and  $4x^3 + 5$  are both polynomial expressions.

In this paper, the focus is on how to help people with visual impairments and blindness to develop their mathematical learning skills by searching and retrieving mathematical documents from the web or on the digital libraries. In this context, a new assistive technology was developed and adapted for braille users to retrieve math expressions. In the second section, we present an overview of existing software solutions for accessibility, studying mathematics and retrieving math expressions. In Section 3, we present mathematical braille codes and converters. In Section 4, we present and discuss our proposed system. The experimental results are in Section 5. Finally, we present the conclusion and future work.

## 2. State of the Art

We review the state-of-the-art process in three stages: we started by performing a keyword search from 2000 to 2017 for phrases like “braille code”; “mathematical expressions retrieval”; “MathML”; “Assistive Technologies”, and many related words. To ensure that we did not miss any important research articles, we conducted the same research on Thomson Reuters (Clarivate), Scopus, and Google Scholar. It was clear that there is considerable lack of this line of research. Second, we reviewed the abstract of 107 articles to keep only papers related to our study. Third, we developed our bibliographic study over three points: accessibility, accessibility to math content, and retrieval of math content.

### 2.1. Accessibility to Mathematical Content by People with Visual Impairments and Blindness

#### 2.1.1. Accessibility

People with visual impairments and blindness, like sighted people, can use computers, tablets, smartphones, and other electronic devices, but differently. They use assistive technologies to help them to easily and independently have access to these devices. Therefore, it is crucial to demonstrate some essential assistance for blind people, encompassing software or hardware systems:

- Screen readers are software applications that allow people with visual impairments to listen to or read the content of a text. It converts the text displayed on the computer screen into speech synthesizer or braille, and the former requires just a simple audio output. However, braille transcription requires additional hardware called refreshable braille displays. In the literature, there are many software solutions, like JAWS (Job Access With Speech), NVDA (Nonvisual Desktop Access), VoiceOver, and Windows Eyes [8].
- A refreshable braille display is an electromechanical system used by people with blindness to read with their fingers like braille hardcopy to decipher the information displayed on the screen. This device can display up to 80 characters on the screen into ephemeral pins, and it changes continuously when the user moves the cursor via control keys. This hardware can also send commands to the computer through the function keys and the routing cursor. With this feature, the blind can perform research, takes notes, and communicate with other people. This device can be used simultaneously or along with a speech-synthesizer system.
- Scanner and Optical Character Recognition allows visually impaired people to scan and read printed books. By connecting this hardware with the computer, the printed document can be scanned and converted into an electronic file that is displayed as text on the computer screen. This text can be read and/or listened to by the blind using refreshable braille displays and screen reader software, respectively. The disadvantage of this material lies in the limitation of digitizing all types of information like images, mathematical notations, and tables.

Braille e-books, note takers, and voice-recognition software are other assistive technologies that help people with visual impairments to independently interact with the computer and sighted people. Thus, they have some limitations in processing and accessing the content of all types of information such as images, tables, mathematical notations; in addition, the price of these assistive devices is high.

### 2.1.2. Assistive Technologies for Braille Users to Study and Access Mathematical Content

People with visual impairments have many barriers while studying and accessing digital Mathematical documents. They need assistive technology to aid with their disabilities. In this section, it is of paramount significance to present some software solutions intended for blind people to independently use and manipulate digital content, especially that related to mathematics.

Braille Math Extension to RoboBraille (BMER) [9] is an automated and flexible system for transcribing schoolbooks created for sighted people into braille books. This system consolidates several braille transcriptions and four math conversion libraries; Sensus SB4 [10] and LibLouis [11] are used for literary braille text conversions, while fMath [12] and UMCL [13] handle the conversion from MathML to a variety of math braille codes.

P. B. Stanley et al. [14] presented an assistive software application for people with visual impairments to use mathematics. All mathematics converted into MathML is translated into Nemeth Braille Code (NBC) [3], and the latter can be directly transferred to a refreshable braille device, downloaded into a braille device, and embossed by a third-party application [15,16].

J. A. Gardner [17] described the new LEAN Math application. He gives a brief overview of the features of this system. LEAN Math is an editor for MathType equations in Microsoft (MS) Word. People with visual impairments can use this application to create, edit, and manipulate math equations in braille and/or audio.

N. Soiffer [18] proposed the latest revision of the MathPlayer4 system [19], which is an Internet Explorer a plug-in that displays MathML and makes it accessible to assistive technology like screen readers and magnifiers. This version includes features to allow the navigation of mathematical expressions. The results of this system indicated that the comprehension rates were similar compared to the preferred modality used by blind students.

A. Salamonczyk and J. B. Pawlowska [20] presented a novel approach to the problem of semantic translation of mathematical expressions to Polish text. To overcome this problem, they used MathML to help and allow blind Polish people to read online documents containing mathematics.

People with blindness can access and study digital mathematics using the aforementioned software solutions. However, they cannot research or retrieve, as independently as sighted people, mathematical notations from digital libraries or the web. These solutions cannot retrieve this kind of information.

### 2.2. Retrieval of Mathematical-Expression Systems

Nowadays, several software solutions have been developed to retrieve mathematical documents from the web or digital libraries. Unfortunately, there is no system for people with visual impairments and blindness. In this section, we present existing systems for math=retrieval information and software applications developed for sighted people.

Math Information Retrieval (MIR) [21] is based on combining math and text keywords in a query and merging the results of relaxed subqueries. The latter are generated from the original query and they consist of  $k$  keywords and  $f$  formulae using the Leave Rightmost Out (LRO) method. The partial result lists of all the subqueries are merged into the final list presented to the user. This research used different querying strategies and result merging. The NTCIR-11 [22] Math Task 2 database was employed to evaluate the strategies in this system.

The development of the Czech Digital Mathematics Library motivated Masaryk University to start developing a Math Information Retrieval system called Math Indexer and Searcher (MIaS). This system is a math-aware full text-based search engine. It supports a combination of text and math

searching. In this subject, Reference [23] presents similarity search for mathematics with different query languages, such as Presentation MathML, Content MathML, Presentation and Content MathML, and Tex. the authors used the NTCIR math task database to evaluate this system.

WikiMirs is a mathematical information-retrieval system for Wikipedia proposed in the literature [24]. It is based upon both textual and structural similarities. This system includes four main modules, specifically, a preprocessor, tokenizer, indexer, and ranker. Experimental results that depend on Wikipedia dataset version 01-10-2012 show that WikiMirs retrieves more accurate results and ranks formulae better in Wikipedia compared with two other search engines, Wikipedia Search and Egomath [25].

MathWebSearch (MWS) presented in the literature [26] is a content-based full-text search engine. It provides low-latency answers to full-text queries that contain formulae and text. MWS combines exact and powerful formula unification with the full-text search capabilities of ElasticSearch for a simultaneous full-text search for mathematical documents. NTCIR Math Task Dataset was used to evaluate the MWS.

A Mathematical Retrieval System proposed in reference [27] used a novel indexing and matching model, and it is based on both textual and spatial similarities. This system includes six components, namely, user interface for WebPages or PDF documents, a preprocessor that converts formulae markups from different sources into Presentation MathML, a tree construction that constructs semioperator trees from MathML, a tokenizer that extracts all terms, an indexer that uses Term Frequency-Inverse Document Frequency (TF-IDF) and level to each term in the inverted index files, and a ranker that finds the matched terms in the index files. The same dataset collected from publicly available WebPages and PDF documents was used to compare the results of this system with MlaS and WikiMirs. This system provides more efficient mathematical-formulae index and retrieval.

The math search presented in Reference [28] proposed a lattice-based approach based on Formal Concept Analysis (FCA). The latter has been used for information retrieval [29–31] as it is a powerful data-analysis technique. This system includes several phases. The first encompasses the conversion of a math query into uniform format as Presentation MathML, then the extraction of features from MathML, after the construction of a mathematical concept lattice with the extracted features, and finally the relevant expressions that are retrieved and ranked when the query and the document share the same attributes. The dataset collected in this paper is 489 math expressions with different functions.

S. Yang et al. [32] presented a math information-retrieval system using a novel data structure called FDS (Formal Description Structure) index. They realized exact matching between a math query and the candidate formula data rather than the document. The query-mode algorithms used in this system are: global query mode, local query mode, and operational query mode. For evaluating the efficiency of this proposed system, they constructed a dataset with 2557 mathematical expressions collected from several mathematics documents, and they compared the results of this system with two search engines, WikiMirs and EgoMath.

T. Schellenberg et al. [33] presented a math information-retrieval system based on layout substitution trees indexing from a LaTeX dataset. They used a TF-IDF technique to rank or retrieve search results. To evaluate the performance and efficiency of this system, the authors used the same document database and search queries in the Zanibbi and Yuan systems. The results ranked by both systems are comparable.

In this paper, like sighted people, we have proposed our assistive technology for people with blindness to help them study, access, and retrieve mathematical expressions from the web and/or digital libraries.



### 3. Mathematical Braille

#### 3.1. Mathematical Braille Codes

Braille is a tactile writing system using raised dots for the visually impaired and those with strong visual disabilities. The system is named after its inventor, the Frenchman Louis Braille (1809–1852) who lost his sight after an accident.

Mathematical braille is diverse and not universal. There are several braille bodes, for instance:

- Mathematical braille notation, used in France.
- Unified Mathematical Code (UMC), practiced in both Latin America and Spain.
- The Nemeth Code, developed by the mathematician Abraham Nemeth, primarily used in the United States and the Middle East.
- The Marburg Code, used in Germany and Austria.
- Mathematical Braille Code, created by the authority of the United Kingdom and used in the United Kingdom and Ireland.

The following table (Table 1) shows the differences between the symbols of mathematical braille codes: the French notation, UMC, and Nemeth.

**Table 1.** Differences between mathematical braille codes: French notation, Unified Mathematical Code (UMC) and Nemeth.


































Symbols	French	UMC	Nemeth
1			
2			
3			
4			
5			
6			
7			
8			
9			
0			
+			

Table 1. Cont.

Symbols	French	UMC	Nemeth
-			
*			
/			
^			
√			
=			
(			
)			
[			
]			
{			
}			

The standard braille with six points allows only few combinations in order to code the characters. To mitigate this problem in computing, they added two points at the end to have eight-point braille. This addition allows making 256 combinations (28), which exactly corresponds to the ASCII table used in computing. In this paper, we use standard braille with six points.

### 3.2. Mathematical Braille Converters

There are numerous braille converters in the literature, like:

- **Labradoor:** (LaTeX-to-braille-door) It converts a LaTeX document including mathematical formulae as well as literary text in braille notation. This project is one of the first attempts at automating the conversion of standard-form mathematical text to print the works in braille, and it is used to produce mathematical texts for secondary-education and university students. Labrador uses only Marburg Braille notation and there is no conversion with other national codes.
- **Liblouis** [11]: It is capable of transcribing numerous document formats (such as DTBook and XML) in braille and supports numerous languages. Liblouis also supports mathematical braille code like Nemeth and Marburg. It is developed as a library to be used in other programs.
- **BraMaNet** [34]: Software that converts Presentation MathML only into French Braille. The transcribed file is a standard textual file so that it can easily be used by any other braille transcription software, such as BraMaNetest based on XSLT technology. Diverse parameters are authorized, including the possibility of modifying the braille output table in order to adapt itself

to any material. It also comes with a script called MetaBraMaNet that automatically converts from a Microsoft Word document containing mathematical formulae written in MathType to braille.

- UMCL [13]: (Universal Maths Conversion Library) It is a software project for the conversion of mathematical formulae between various formats: MathML specific notation and braille.
- Infty [35]: This is an integrated system aiming to make printed scientific documents, including mathematical formulae, accessible. The system consists of three applications' components: an OCR system named "InftyReader", a publisher named "InftyEditor", and conversion tools in various formats. The vocal interface ChattyInfty [36] allows partially sighted users to access and publish the expressions with a speech output.
- Math2Braille: open-source software that allows the conversion of MathML files into braille. The process of converting MathML into braille relies on protocols and procedures that were developed in a previous project on access to music.

#### 4. Retrieval Mathematical Equations System for People with Visual Impairments and Blindness

Due to the complexity of our system, we have taken a bottom-up approach to better represent our work. We start by drawing up the overall schema of our system and we dissect its components.

Figure 1 illustrates the general architecture of our proposed system. The system allows people with visual impairments and blindness to semantically search the mathematical equations from the web or digital libraries. The development of this architecture is inspired from classification systems [37].

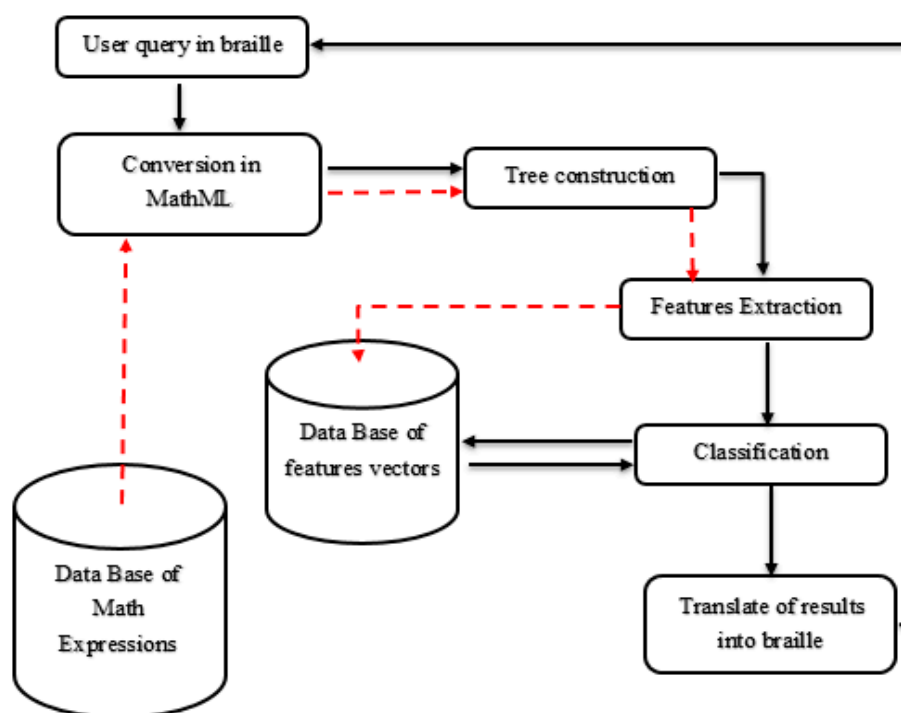


Figure 1. Proposed system.

The system takes place in two modes, specifically, offline mode or indexing mode described by the discontinued line, and online mode described by arrows. In offline mode, the mathematical equations of the database are converted into MathML [6] code. Afterward, there comes the semantic analysis of the equations in presentation MathML and the construction of the tree substructure for every formula, and then the extraction of the characteristics through this tree substructure in the form of features vectors. Finally, these vectors are stored in the database.

In online mode, the user launches a request equation in braille code. The request is firstly pretreated in MathML using a presentation markup that analyzes it semantically and then, based on

that, builds the tree substructure of the request and the extraction of characteristics in the form of the features vectors. Finally, there is the classification of similar vectors to the request in the database. The results of the classification are converted into braille and presented to the user. We explain in detail the various phases of the system in the following sections.

#### 4.1. Transcription of Braille Expression into MathML Code

The difficulties of understanding a braille code lead us to use the Accessible Equation Editor (AEE) [38] extension to edit the equations in Nemeth Braille Code. With this extension, we can transcribe equations into braille, which is used as a request equation of our system. The query conversion interface is shown in Figure 2.

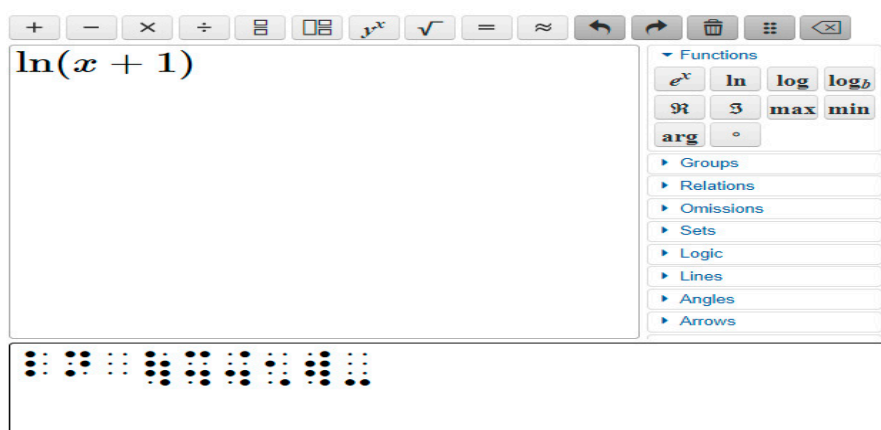


Figure 2. Braille conversion interface.

The preprocessing of braille queries into uniform mathematical representation plays an important role in the field of mathematical research. There are several markup languages like LaTeX, and presentation and content MathML. MathML is recommended by W3C and it is the most common format for web presentation. Currently, it is used by numerous systems to retrieve mathematical expressions on the web [27,28] and by other applications, such as a MathML to Nemeth translator [14].

In our system, we use presentation MathML to code the semantic and notational structure of mathematical equations (Figure 3). This formula representation is marked and limited by two markups, <math> and </math>.

```

1 <math display='block' xmlns='http://www.w3.org/1998/Math/MathML'>
2   <mrow>
3     <mi>log</mi><mo>(</mo>
4       <mi>sin</mi><mo>(</mo><mfrac>
5         <mrow>
6           <mi>x</mi><mo>+</mo><mn>1</mn></mrow>
7           <mn>3</mn>
8         </mfrac>
9       <mo>)</mo><mo>)</mo>
10   </mrow>
11 </math>

```

Figure 3. MathML code of mathematical expression  $\log(\sin(\frac{x+1}{3}))$ .

#### 4.2. Semantic-Tree Construction

The structures of the mathematical formulae are naturally hierarchical. They are read and understood in a hierarchical way. The structure of the upper level is more important than those of the lower levels. The upper level generally determines the category and skeleton of the mathematical formula, while the knots of the lower levels usually keep variables or constants that do not bring sense to the content of the formula structure.

The structure representation tree can be directly extracted from the MathML code (Figure 4). This representation loses numerous useful semantic contents in searching for mathematical formulae.

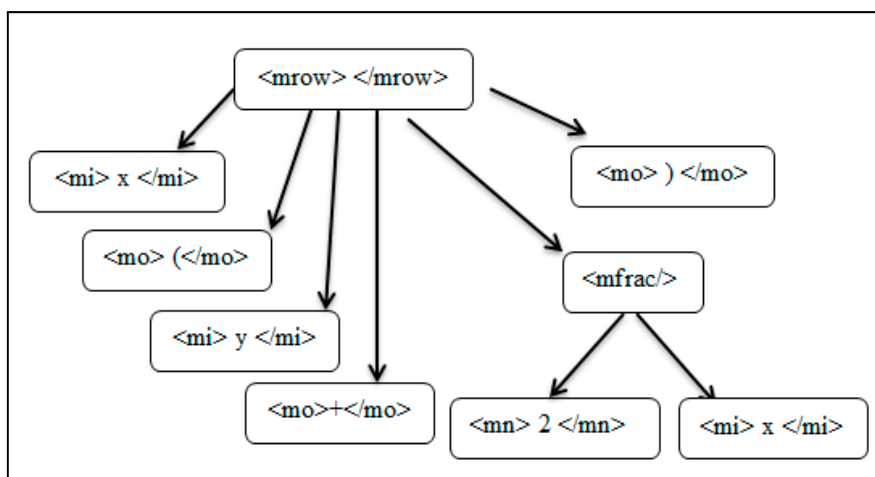


Figure 4. Structure representation tree in MathML of equation  $x(y + \frac{2}{x})$ .

This representation can have two types of relations:

- One-dimensional: Symbols are horizontally connected, e.g., expressions connected by operators '+', '-', '\*', etc. Symbols of this type of relation are connected by the same horizontal label, `<mrow>`, in MathML.
- Bidimensional: Symbols are connected in nonlinear relations. For example,  $\sqrt{\quad}$ ,  $\sum$ ,  $\int$ ,  $\prod$ , etc.

Structures of the one-dimensional relations cannot be used in mathematical research. and for that reason, semantic interpretation is necessary in the conversion of the structure representation tree in one-dimensional relation into a matching semantic presentation. This conversion uses methods and algorithms that interpret the operator and level priority (Figure 5).

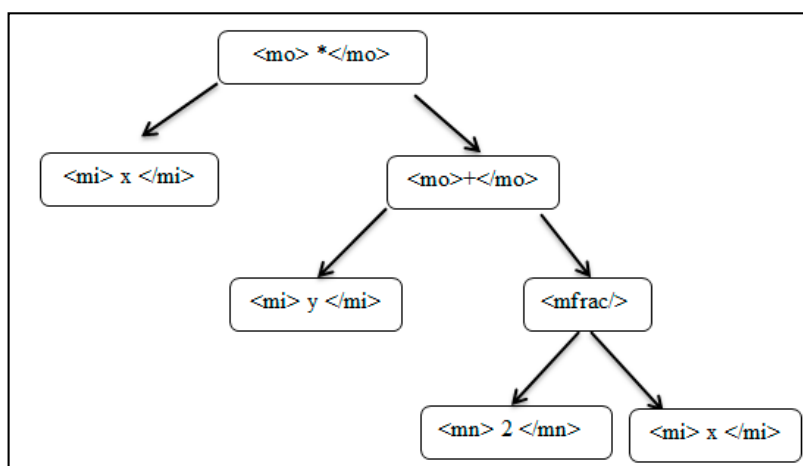


Figure 5. Semantic tree in MathML of equation  $x(y + \frac{2}{x})$ .



Our algorithm is based on the definition of different levels for each mathematical formula. The first level is defined by the search of all the main operators (+, −, \* ... ) identified in MathML<mrow>, linking them by parenthesis and/or functions (trigonometric, logarithmic, exponential). Everything between a parenthesis and a function's arguments is replaced by the term «exp». The values of these «exp» are stocked to be used in the next level.

For example:  $\sqrt{x+1} + x^2 \ln(x^2 + 1)$

Becomes  $\sqrt{\text{exp}} + \text{exp} \ln(\text{exp})$

Other levels are defined by repeating the same procedure for every «exp» term.

After the recursive execution of this method from the lowest to the highest level of the semantic tree extracted from the MathML code of an equation, a semantic structure in the form of a tree of several levels is built. Figure 6 shows this method with the equation example  $\cos(x) * \sin(\ln(x))$

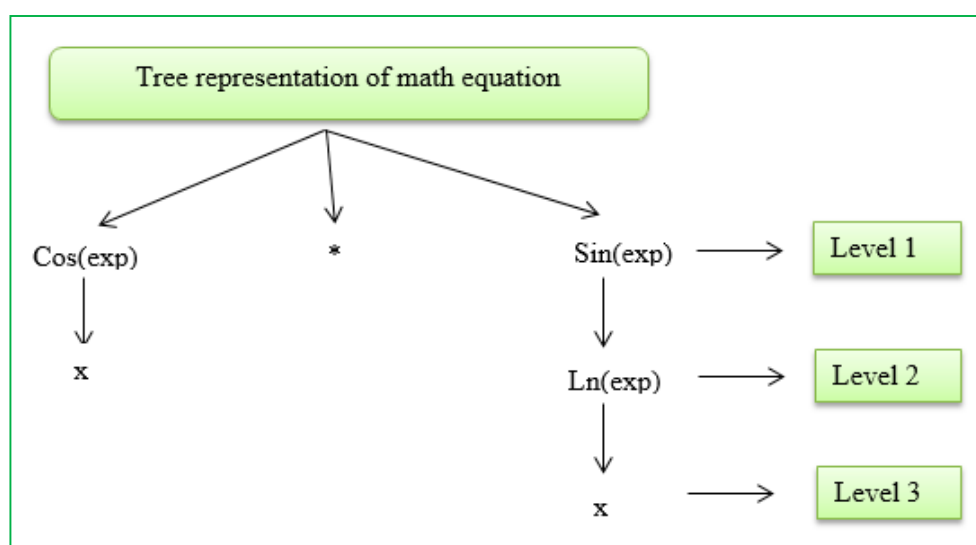


Figure 6. Multilevel tree construction of equation  $\cos(x) * \sin(\ln(x))$ .

#### 4.3. Feature Extraction

The extraction phase aims to extract the features vector of every mathematical formula of the database. For that, we used the conventional vector model with the multilevel extraction method: Extraction from the multilevel tree shown in Figure 6 for every mathematical formula in the database.

The conventional vector model is based on the vector concept. This model introduces relevance weights associated to features terms by using different matching functions. It assumes that each equation can be presented by a vector whose nonzero coordinates correspond to the equation's feature terms. The search for similarity between two equations is executed by calculating the similarity between vectors associated with each of them.

The extraction method's base concepts become as follows:

$E$ : Database set of the equations  $E_1, E_2, E_3, \dots, E_n$ .

$T$ : Set of features terms of a database's equations. The terms are the feature-vector variables: operators ('+', '−', '\*' ...) and functions ( $\cos$ ,  $\sin$ ,  $\sqrt{\phantom{x}}$ ,  $\log$ , ...).

$V(T)$ : Space vector of dimension  $|T| = n \in \mathbb{N}$  is defined as the number of terms contained in universe  $T$ .

$V_{il}$ : Features vector of equation  $E_i$  in an  $l$  level of the multilevel tree.

$W_{ikl}$ : Value of the relevance weight of feature term  $T_k$  in equation  $E_i$  in the  $l$  level of the multilevel tree. In our case, if  $T_k$  does not belong to the descriptor terms of equation  $E_{il}$ ,  $W_{ikl}$  takes a value of 0; otherwise,  $W_{ikl}$  takes a value of 1.

The application of this method for every level of the multilevel tree allows to have for each equation  $E_i$  a set of features vector  $V_i$  ( $V_{i1}, V_{i2}, \dots, V_{in}$ ). The first level's vector  $V_{i1}$  is obtained by calculating the occurrence number of each term  $T_k$  of vector space  $V(T)$  in equation  $E_i$  while ignoring the content of the terms “exp” and stocking them in features vector  $V_{i1}$ . The “exp” terms are treated by repeating the same method applied at the first level. We repeat the same procedure for all remaining levels of equation  $E_i$ . The variable and constant terms contain the number of occurrences of the different variables and constants that exist in each level in equation  $E_i$ .

#### 4.4. Classification

When the user launches a request, the system proceeds by converting it into MathML code and, subsequently, the multilevel tree is constructed. After that comes the extraction of the features using the multilevel method and then storing the extracted features in the form of vector  $V_Q$ , where  $V_Q(V_{Q1}, V_{Q2}, \dots, V_{Qn})$ . The search for similarities between query and database equations is performed by the distance calculation. Therefore, we proposed in this paper to use the K-Nearest-Neighbor (KNN) method.

The KNN algorithm is a widely used classification technique [39–41]. It is one of the simplest algorithms as it stocks all available cases and classifies the new ones by similarity measures. Even with such simplicity, KNN can output highly competitive results. We have used this algorithm to look for similarities between the mathematical equation of the request and that of the database. This similarity is based on the calculation of the distance between features vector  $V_Q$  where  $V_Q(V_{Q1}, V_{Q2}, \dots, V_{Qn})$  stemming from the extraction method. The distance formulae can be expressed as follows:

- Euclidian distance:

$$d(V_{Ql}, V_{yl}) = \sqrt{\sum_{i=1}^T (V_{Qli} - V_{yli})^2} \quad (1)$$

where:  $V_{Ql}$  and  $V_{yl}$  are the features vectors of the request equation and an equation  $E_y$ , respectively, for level  $l$ .

$T$ : is the number of terms in vector space  $V(T)$ .

- Minkowski distance:

Euclidean distance is a special case for  $p = 2$  of Minkowski distance.

$$d(V_{Ql}, V_{yl}) = \left( \sum_{i=1}^T |V_{Qli} - V_{yli}|^p \right)^{\frac{1}{p}} \quad (2)$$

The calculation of the similarity using the multilevel method is based on multilevel structural and semantic similarity between mathematical equations. The degree of similarity is obtained according to all levels of the multilevel tree (Figure 6). After having defined the levels of the request and the features vector  $V_Q(V_{Q1}, V_{Q2}, \dots, V_{Qn})$ , as a first step, we search all mathematical expressions having similar features vectors of the first level  $V_{i1}$  to  $V_{Q1}$  by the calculation of the distance between them. Then, we proceed to the next level just for the similar equations retrieved in the first level. This procedure is repeated in the second level. We continue with the same technique for all levels of the query until we obtain the vectors closest to the request. Reverse indexing allows us to retrieve all mathematical equations corresponding to similar founded feature vectors.

#### 4.5. Braille Transcription

In this step, we used Math2Braille in order to convert similar query results into braille. As we do not have a refreshable braille display to examine the output, the latter is displayed on the screen. Figure 7 illustrates the system output.

#	Equation	
	$x + 1$	
	<code>&lt;math xmlns="http://www.w3.org/1998/Math/MathML" fontsize="36px"&gt;&lt;mrow&gt;&lt;mi&gt;x&lt;/mi&gt;&lt;mo&gt;+&lt;/mo&gt;&lt;mn&gt;1&lt;/mn&gt;&lt;/mrow&gt;&lt;/math&gt;</code>	
1	$1 + x$	
2	$1 + xy$	
3	$zx + 5$	
4	$\frac{1}{x+1}$	
5	$\frac{x+1}{x-1}$	

Figure 7. Visualization the system output in braille.

## 5. Experimental Results

### 5.1. Database

Analyzing the performance of a retrieval mathematical-expression system is not an easy task, because there are few standard reference datasets, contrary to other more common tasks of retrieval information. Normally, each retrieval mathematical-information system builds its own database for evaluation. The comparison between our system and others is also difficult since they are either unavailable or inaccessible.

We created a data system composed of mathematical expressions to evaluate the efficiency of the suggested system. All of the data are constructed using MathType, an interactive tool for the creation of mathematical material found in Microsoft Word and PowerPoint under the MathType Rubin tab to facilitate editing, inserting, and creating mathematical equations. The game elements of the dataset can be easily converted into MathML or LaTeX (Figure 8).

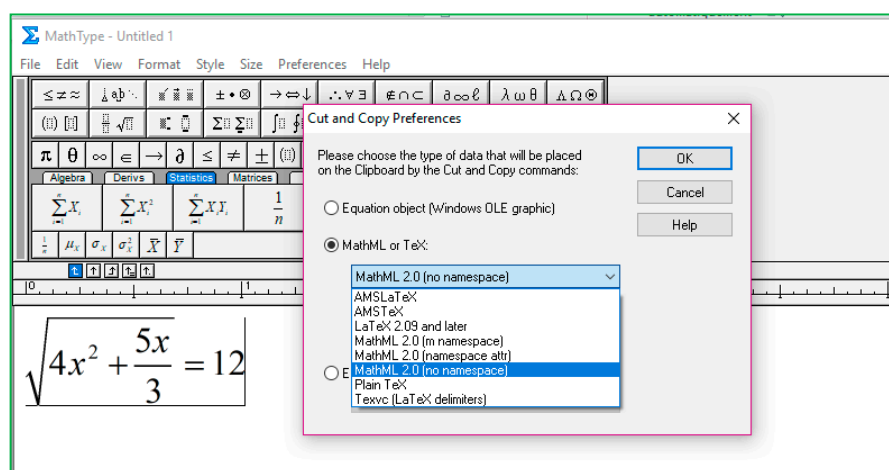


Figure 8. Conversion of mathematical equation into MathML or LaTeX.

In this set, we created 6925 mathematical expressions using symbols from five languages: Latin, Arabic, Tifinagh, Hebrew, and Japanese. For each language, we wrote 1385 different types of mathematical expressions, such as polynomial, algebraic, statistical, trigonometric, and logarithmic expressions. In this paper, we used a Latin set to evaluate the performance of our system.

### 5.2. System Results

In this section, we present the results obtained from the suggested system using the KNN algorithm with Minkowski distance.

The query results of the proposed system are shown in Table 2 below. For each request, we present the first six results in braille. To make it easier to read braille, we converted the queries and results into readable format.

**Table 2.** Obtained results.

Queries	Queries in Braille	Results		
$\ln(x)$	$\ln(x)$	$\ln(x) + x$	$\ln(a + b)$	$\ln(y)$
		$\ln(x) + 7$	$\ln(1 - x)$	$1 - \ln(\sqrt{a} + 2)$
$\sin(x)$	$\sin(x)$	$\sin(x)$	$x\sin(x)$	$\sin(5 + x)$
		$\sin(2) + \sin(1 + b)$	$\int \sin(x + 2)dx$	$\sin(5 + x^2)$
$1 + x$	$1 + x$	$1 + xy$	$zx + 5$	$\frac{1}{x} + y$
		$\frac{1}{x+1}$	$e(2) + 1$	$e(a + \frac{a}{2})$
$\frac{1}{x}$	$\frac{1}{x}$	$\frac{1}{x} + y$	$\frac{1}{x+1}$	$\frac{1+x}{\tan(x)}$
		$\frac{2x}{1-x}$	$\frac{x+1}{x-1}$	$\frac{\ln(x)+1}{x}$
$\ln(\sin(1+x))$	$\ln(\sin(1+x))$	$2\ln(x)\ln(y) = 1$	$1 - \ln(\sqrt{a} + 2)$	$\ln(a) - \ln(a + b)$
		$x\sin(x)$	$\ln(1 - x)$	$\ln(x) + \ln(x + y)$
$e^{(\cos(y)+\sin(y))}$	$e^{(\cos(y)+\sin(y))}$	$\frac{1+\sin(x)}{1-x}$	$e(2) - 1$	$\frac{1-\cos(x)}{1-x}$
		$e(2) + 1$	$\int \sin(x + 2)dx$	$\sin(e(x + 1))$

### 5.3. Discussions

Braille users face many difficulties to study and access to the content of scientific documents containing mathematical notations. With this new algorithm proposed, the people with visual impairments can independently retrieve mathematical expressions and develop their mathematical learning.

In Table 2, we found that:

The obtained results are several notations, for example, for the test request  $\ln(x)$  we found  $\ln(y)$  and  $\ln(a + b)$ . These results show that our system can retrieve similar math equations with several notations. The suggested system takes into account the structural and semantic aspect to retrieve math expressions. This characteristic helps learners with visual





6. Ausbrooks, R.; Buswell, S.; Dalmas, S.; Devitt, S.; Diaz, A.; Hunter, R.; Smith, B.; Soiffer, N.; Sutor, R.; Watt, S. Mathematical Markup Language. (MathML) version 2.0. 2000. Available online: <https://www.w3.org/Math/> (accessed on 23 October 2018).
7. Asebriy, Z.; Raghay, S.; Bencharef, O.; Kaloun, S. A semantic approach for mathematical expression retrieval. *IJACSA* **2016**, *7*, 190–194. [[CrossRef](#)]
8. Wongkia, W.; Naruedomkul, K.; Cercone, N. i-Math: Automatic math reader for Thai blind and visually impaired students. *Int. J. Comput. Math. Appl.* **2012**, *64*, 2128–2140. [[CrossRef](#)]
9. Vlad, P.C.; Tanja, S.; Lars, B.C. Braille Math Extension to RoboBraille: A Universal Software Solution for Converting Math into Braille. In Proceedings of the ICCHP 2016, Linz, Austria, 13–15 July 2016; pp. 15–18.
10. Christensen, L.B.; Klaus, J. Multilingual two-way braille translation. In Proceedings of the ICCHP '96 Proceedings of the 5th International Conference on Computers Helping People with Special Needs, Linz, Austria, 17–19 July 1996.
11. LibLouis. Available online: <http://www.liblouis.org> (accessed on 1 October 2018).
12. fMath. Available online: <http://www.fmath.info> (accessed on 28 September 2018).
13. UMCL. Available online: <https://sourceforge.net/projects/umcl> (accessed on 23 October 2018).
14. Stanley, P.B.; Karshmer, A.I. Translating MathML into Nemeth Braille Code. In Proceedings of the 10th International Conference, ICCP 2006, Linz, Austria, 11–13 July 2006; pp. 1175–1182.
15. Duxbury Systems. Available online: <http://www.duxburysystems.com/> (accessed on 27 September 2018).
16. Amaya XML Editor. Available online: <https://www.w3.org/Amaya/> (accessed on 20 October 2018).
17. Gardner, J.A. The Lean Math Accessible MathML Editor. In Proceedings of the ICCHP 2014, Paris, France, 9–11 July 2014; Part I, LNCS 8547; pp. 580–587.
18. Soiffer, N. A Study of Speech Versus Braille and Large Print of Mathematical Expressions. In Proceedings of the ICCHP 2016, Linz, Austria, 13–15 July 2016; Part I, LNCS 9758; pp. 59–66.
19. Soiffer, N. Browser-independent Accessibilité Math. In Proceedings of the W4A 15, Florence, Italy, 18–20 May 2015.
20. Salamonczyk, A.; Pawlowska, J.B. Translating of MathML Formulas to Polish Text, Example Applications in Teaching the Blind. In Proceedings of the 2nd International Conference on Cybernetics (CYBCONF), Gdynia, Poland, 24–26 June 2015.
21. Liska, M.; Sojka, P.; Ruzicka, M. Combining text and Formula Queries in Math Information Retrieval. In Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM 2015), New York, NY, USA, 19–23 October 2015.
22. Aizawa, A.; Kohlhase, M.; Ounis, I.; Schubotz, M. NTCIR-11 Math-2 Task Overview. In Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo, Japan, 9–12 December 2014; pp. 88–98.
23. Liska, M.; Sojka, P.; Ruzicka, M. Similarity Search for Mathematics: Masaryk University team at the NTCIR-10 Math Task. In Proceedings of the 10th NTCIR Conference, Tokyo, Japan, 18–21 June 2013.
24. Hu, X.; Gao, L.; Lin, X.; Baker, J.B. WikiMirs: A Mathematical Information Retrieval System for Wikipedia. In Proceedings of the 13th ACM/IEEE-Cs Joint Conference on Digital Libraries ACM, Indianapolis, IN, USA, 22–26 July 2013; pp. 11–20.
25. Egomath. Available online: <http://egomath.projekty.ms.mff.cuni.cz/> (accessed on 18 June 2018).
26. Hambasan, R.; Kohlhase, M.; Prodescu, C. MathWebSearch at NTCIR-11. In Proceedings of the 11th NTCIR Conference, Tokyo, Japan, 9–12 December 2014.
27. Lin, X.; Gao, L.; Hu, X.; Tang, Z.; Xiao, Y.; Liu, X. A mathematical retrieval system for formulae in layout presentations. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR 14), Gold Coast, QLD, Australia, 6–11 July 2014; pp. 697–706.
28. Nguyen, T.T.; Cheung, S.H.; Chang, K. A lattice-based approach for mathematical search using Formal Concept Analysis. *Int. J. Expert Syst. Appl.* **2012**, *39*, 5820–5828. [[CrossRef](#)]
29. Carpineto, C.; Romano, G. Order-theoretical ranking. *Int. J. Am. So. Inf. Sci.* **2000**, *51*, 587–601. [[CrossRef](#)]
30. Cheung, K.S.K.; Vogel, D. Complexity reduction in lattice-based information retrieval. *Int. J. Inf. Retr.* **2005**, *8*, 285–299. [[CrossRef](#)]
31. Hao, S.; Shi, C.; Niu, Z.; Cao, L. Concept coupling learning for improving concept lattice-based document retrieval. *Int. J. Eng. Appl. Artif. Intell.* **2018**, *69*, 65–75. [[CrossRef](#)]

32. Yang, S.Q.; Tian, X.D.; Yu, B.T. A maintenance algorithm of FDS based mathematical expression index. In Proceedings of the International Conference on machine learning and Cybernetics, Lanzhou, China, 13–16 July 2014.
33. Schellenberg, T.; Yuan, B.; Zanibbi, R. Layout-Based Substitution Tree Indexing and Retrieval for Mathematical Expressions. In Proceedings of the Document Recognition and Retrieval XIX, Burlingame, CA, USA, 22–26 January 2012.
34. BraMaNet. Available online: <http://handy.univ-lyon1.fr/MH/bramanet/bramanet.php> (accessed on 24 June 2018).
35. Suzuki, M.; Kanahori, T.; Ohtake, N.; Yamaguchi, K. An Integrated OCR Software for Mathematical Documents and Its Output with Accessibility. In Proceedings of the International Conference on Computers for Handicapped Persons, Paris, France, 7–9 July 2004; Springer: Berlin/Heidelberg, Germany, 2004; pp. 648–655.
36. Komada, T.; Yamaguchi, K.; Kawane, F.; Suzuki, M. New Environment for Visually Disabled Students to Access Scientific Information by Combining Speech Interface and Tactile Graphics. In Proceedings of the International Conference on Computers for Handicapped Persons, Linz, Austria, 11–13 July 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1183–1190.
37. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2000; p. 680.
38. Accessible Equation Editor. Available online: <http://accessibility.pearson.com/aee/> (accessed on 7 July 2018).
39. Maillo, J.; Ramírez, S.; Herrera, I.T.F. kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data. *Int. J. Knowl.-Based Syst.* **2017**, *117*, 3–15. [CrossRef]
40. Pedronette, D.C.G.; Gonçalves, F.M.F.; Guilherme, I.R. Unsupervised manifold learning through reciprocal kNN graph and Connected Components for image retrieval tasks. *Int. J. Pattern Recogn.* **2018**, *75*, 161–174. [CrossRef]
41. Zhan, Y.; Liu, J.; Gou, J.; Wang, M. A video semantic detection method based on locality-sensitive discriminant sparse representation and weighted KNN. *Int. J. Visual Commun. Image Represent.* **2016**, *41*, 65–73. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).