A transcription tool for mathematical Braille

Moço Victor and Archambault Dominique
INSERM U483 / INOVA

Keywords: Mathematical Braille, automatic translator, central representation

Abstract : In the framework of the Vickie project, funded by the European Commission, which aim is to provide tools faciliting inclusive education for visually impaired children, we have developed a tool allowing automatic transcription of Mathematical formulas. This tool is based on a central representation of the formula, in order to separate input from output, that facilitate the integration of new formats (the mathematics Braille notation is not international). Indeed for each format we need a parser to create this representation and an output module to generate a formula from the internal representation. A prototype was developed in java that includes major features and is able to process mathematical contents up to the middle of secondary school. Latex, MathML and French Mathematical Braille are supported. It also allows the user to select a Braille table. A website allows to use this prototype.

1.Introduction

1.1. The Vickie Project

Vickie is an acronym for **V**isually **I**mpaired **C**hildren **K**it for **I**nclusive **E**ducation. This European project must produce all the necessary tools for the integration of impaired children in mainstream school. In schools, Learning is based on writing. Blind student use a different writing mode : the Braille. To preserve the interaction between the sighted (professor or another children) and the impaired children, we need - in other things -to make in few time the translation "sighted document" to Braille document and inversely. To facilitate translations, the sighted person and the blind student use automatics translators. In this study, we focus on the translation of digital mathematical documents.

1.2. Digital mathematical Document

Blind student can read digital document with refreshable Braille display. The refreshable Braille display use Braille table to make a relation between digital characters and Braille characters. Remind that a Braille 6-dots character is a matrix three rows and two columns, we can write 64 differents characters. A 8-dots Braille exists (digital Braille: four rows and two columns) but, in our case, we only use the 6-dots Braille. The Braille code has special characters -escape characters- which change the meaning of the next characters. These escape characters, we allows to create all the mathematics characters.

Braille code and "digital" writing - as normal text- is an one dimensional writing. But, mathematics use two-dimensional writing (in case of using fraction or square root for instance). The mathematical Braille or digital mathematical document must simulate the two-dimensional writing using specific rules. In this study, we don't translate digital document when the mathematic formula is a picture. In this case, we need to use the optical character recognition technology. In our case, we must change the specific rules used by the document to create an another digital document. This new document can be printed.

2.Central representation

2.1. Automatic translator for mathematical Braille

There are several automatic translator for the mathematical Braille. For example, there are :

<u>BraMaNet</u> [1] (Braille Mathématique sur Internet) : it translates a formula write in MathML [2] presentation tag in French mathematical Braille notation.

<u>LaBradoor</u> [3] (Latex to Braille Door) : it translates a formula written in Latex in the German mathematical Braille (Marburg notation).

<u>M.A.V.I.S[1] project</u> [4] : has a translator which translates a formula written in Latex in the American mathematical Braille (Nemeth code).

These translators have been developed in the goal to produce mathematical documents for Blind students. As you can see, these translators use specific format for the "sighted document" and produce a specific mathematical Braille notation. For our translator, we decided to use a central representation of the mathematical formula.

2.2. Why a central representation ?

Contrarily to the mathematical mainstream notation the mathematical Braille notation is not an international notation (Marburg notation, Nemeth notation, ...). To translate correctly a mathematical formula, we need to know the notation whose employed to use the good rules of translation. The Vickie project was an European project, consequently, we need to translate in several Braille notation. Accepting several notation for the mathematical Braille, we have decided to use two formats (Latex and MathML) for the sighted document. We have choosen Latex because it's an usefull format used in universities and MathML because it's the standard to wrote mathematics in Internet.

Traditionally when we make a translation, we have a dictionary which makes the relation between the "word" of the input format and the "word" of the output formats. In our case, we have a lot of formats (we can use one format in input or in output), by consequence, we must produce a lot of dictionaries. Moreover, the different format use different rules to write the same formula. For instance, the formula 1+2=3 is :

1+2=3 in Latex

= + 1 2 3 in MathML semantic's version[2]

But, the semantic of the mathematical formula does not depend on the format. For this reasons, we have decided to use a central representation of the mathematical formula. This central representation is a semantic tree which show the semantic structure of the formula. Reading this tree with specific rules, allow to write the formula in the output format. This tree is built with the Polish notation of the formula. The Polish notation puts the operator before these parameters. It's the notation used by the MathML semantic's version.

2.3. Central language

To be interesting, the central representation of the formula must be written independently on input or output format. We decided to write our central representation of

---

[1]M.A.V.I.S : Mathematics Accessible to Visually Impaired Student.

[2]To simplify the formula, we don't write the MathML tag but only the structure of the formula.

the formula using a central language. This central language is important because it allows the separation between the input and the output format. We make only one dictionary by format : mathematical format (Latex, mathematical French Braille ...) / central language.

As all languages, we must use words of different nature. We use a « sample » data like letter and we use complex structure like square root for instance. For us, a complex structure is a structure which have a begin tag and an end tag. In general, these complex structures change with the formats we use. To know the different nature of words, we use a specific dictionary. This dictionary gives a nature at every word of the central language.

3.Specification of the translator.

To translate a document, we need to extract the semantic of the document. Then, we write this semantic in a new document with new rules (in general). Our translator structure follow these steps. First, we build the central representation using the central language. Then, we use -read - this central representation to write the same formula in the output format. The building module depends on the input format rules, and the reading module depends on the output format rules.

3.1. Building module

The building module must extract the semantic of the formula. This operation is not easy and depends on the input format. For example, if we have the sample formula : « 2a ». Everybody knows that it is the multiplication between the number 2 and the letter a (Remind that the translator uses a dictionary to know the nature of a word).

Latex and the French mathematical Braille write exactly this formula. Their building module need to know the rule : « when you have a letter following a number then it is a multiplication ». But, in MathML presentation tag, you must write the implicit multiplication using the entity « ⁢ »(You can write only « 2a » but you don't follow the MathML specification). In this case, you don't need to use a previous rule. The building module must build differently the explicit and the implicit multiplication. In the first, we write the mathematical symbol but not in the second.

There are formulas for which meaning extract is difficult. For instance : « a(x+1) » can be :

- the multiplication between the letter a and the complex structure (x+1) (1)

or

- the function a() with the parameter x+1. (2)

The MathML can make the difference between the two formulas. But, Latex and the French mathematical Braille don't make the difference. In this case, for the moment, we always translate the fomula as a multiplication. It is not always semantically correct (for the MathML), but « visually » and in Braille, the translation is correct.

In conclusion, to build correctly the central representation, we need to know :

- general rules, like the implicit multiplication
- all the specification of the input language
- and to have all the words and nature of the input language.

For the mathematical Braille, we must do a more step. This step must be located before to build the central representation. In the introduction, we wrote that the refreshable Braille Display use a Braille table to make a relation between the Braille symbol and the digital character. But, several Braille tables exits. Blind people can make his own Braille table. To translate correctly a formula, it's necessary to know the table used to write the formula. There exists several methods to write a Braille symbol without a Braille table. Before the building module, we must translate the Braille formula into a formula independent from Braille table. This operation is not a semantic operation ; it only changes the character written.

3.2.Reading module

The reading module takes the central representation and writes the formula in the output format. This module is more simple that the building module. In this step, we only read the central representation. We only follow the rules of the output format to write correctly the answer. To make correctly this step, it's important to know :

- all the specification of the output format.
- To have all the words of the central language.

It's necessary to have all the words of the central language because, in this case we are sure that all the words of the central representation have a translation.
We use the same dictionaries for the same format (in input or in output). But, there are several words in the central language which were used for the output only. These words don't exist on the mathematical format.
Like in the building module, we must do an another translation if we use the Braille as the output format. This translation write the answer with a specific Braille table.

The building and the reading module are independent. We can have one of these module without the second. For instance, with our prototype, you can take the « old » French mathematical Braille (before September 2001) in output but not in input.

In general, Braille formula are longer than the sighted formula. The blind student can't « see » the formula in its totality. It's important for the Blind student to have a formula with unuseful Braille symbol. For this reason, the reading module analyzes the mathematical formula and writes only the necessary Braille symbols.

4.Current Statement

The translator was developed in Java. It supports Latex, MathML presentation tag and the French Braille notation. It will be included on software service for the vickie project. We have developed a simple interface for demo purpose. A site which allow stesting the translator was also create. You can find this site at the following address :

http://inova.snv.jussieu.fr/maths.

It allows users to choose the Braille table and the size of the refreshable Braille display for the output. The users can see the graphical formula if he uses a MathML compatible browser. And, if he chooses the braille as the output format, he can see the formula with the graphical Braille symbol.

5.Conclusion

Our translator is based on the central representation with a central language. The dictionaries are very important for the translation. The translator is under development, for the moment, we don't translate all the mathematical formulas. We can translate algebrique formulas (indice, root, fraction ...). We don't have all the mathematical symbols.

The input and the output format are independent. If you want, you can make Braille to Braille translation or sighted mathematical notation to sighted mathematical notation. The translator doesn't compare the input and the output format.

Acknowledgement :

The Vickie project is funded by the European Commission[3], on the IST (**I**nformation **S**ociety **T**echnologies) Programme 2001 (FP5-IST-Systems and Services for Citizen/persons with special needs), under the references IST-2001-32678.

References :

[1] Schewbel F, (2002), retrieved April 16 2003 from http://handy.univ-lyon1.fr/projets/bramanet/

[2] MathML (Mathematic Markup language) : http://www.w3.org/Math

[3] Miesenberger K., Batusic M., Stoëger B., (1998), retrieved April 16, 2003 from http://www.jussieu.fr/inova/publi/ntevh/labradoo.htm

[4] Karshmer A. I. (2002), *Access to mathematics for Blind students* retrieved April 16, 2003 from http://www.snv.jussieu.fr/inova/villette2002/res5.htm

[3]The contents of this paper is the sole responsability of the authors and in no way represents the view of the European Commission or its services.