MathML for the Management of Mathematical Formula in Text Editor

Vo Trung Hung
University of Danang, Danang,
Vietnam

Cao Xuan Tuan
Ministry of Education and Training,
Hanoi, Vietnam

Abstract— **In this paper, we present the results of research on the application of MathML and Amaya open source to build an environment as an editor for the document and web page content. The highlight of this editor is that users can easily edit, display, copy and search a mathematical formula. We propose using the MathML standard to store content of mathematical formulas, to build functions, and add them into Amaya open source to make the copying, searching for mathematical formulas. When we copy a mathematical formula from an editor to current environment (or contrary), we implemented the handling (format transformation) in the buffer when copying.**

Keywords—MathML; mathematical formula; content management; editor; standard

I. INTRODUCTION

MathML is not yet widely used, but in the future this markup language will become popular for display the contents of Web or documents which are contained mathematical formulas. The proposal question is that the support of browsers and how to edit the formula in MathML format.

Currently, there are a lot of software that allows editing of mathematical formulas such as MS Word, OpenOffice.org Writer, Acrobat, LaTex... However, every editor software is used a different standard for content of mathematical formula. This makes difficult when copy formulas between applications. Especially, we can't perform a search request for mathematical formula on the current editing software. Therefore, the research on how to store, edit, search the mathematical formula is needed; there are significant scientific and practical high.

The main object of our research is to find a best standard to store the content, develop an editor tool, display and search mathematical formulas. The ultimate goal of our proposed solution is to copy directly mathematical formulas from an editor to other one.

For storage standard, we suggest using the standard MathML because this is supported by the majority of software text editor and especially supported by the web browser. To develop software text editor, we use open source Amaya and integrated tools that we developed for search and copy mathematical formulas (through actions on the keyboard).

In this paper, we present the results of research overview concerning the management of mathematic formula on computer, the tool supports mathematical formula editor and MathML markup language. In the next section, we present the application, the general process model, and proposed solutions to the built an editor environment which is integrated tools for searching, copying mathematical formulas.

II. RELATED WORKS

A. Editors supported for mathematical formula

A formula editor is a name for a computer program that is used to typeset mathematical works or formula.

Formula editors typically serve two purposes: the fist, they allow word processing and publication of technical content either for print publication, or to generate raster images for web pages or screen presentations. The second, they provide a means for users to specify input to computational systems that is easier to read and check than plain text input and output from computational systems that is easy to understand or ready for publication.

- **TeX** is a typesetting system developed by Donald E.Knuth in 1970 [2]. TeX is popular in academia, especially in mathematics, computer science, economics, engineering, physics, statistics, and quantitative psychology.
- **LaTeX** is a free text editor [4]. LaTeX includes a script that allows users to edit and print documents with high quality. LaTeX asks the editor to define the logical structure of the text through a series of commands to be installed and in writing. It will then compile to file.tex on text file format.
- **HTML (HyperText Markup Language)** is a markup language designed for sites dedicated to agree on the format and presentation of documents on the World Wide Web [6]. Although we had to take the standard mathematical formulas into HTML documents but most browsers are not supported or lack the necessary fonts. Moreover, on some browsers that support mathematical formulas, we will see equation is shown unclear and particularly difficult to express a "natural" formula of users.
- **MathML (Mathematical Markup Language)** is a markup language and it can specify mathematics in Web pages as text. The structure of MathML is not as TeX or LaTeX but can be easily used by the browser and can display clearly the mathematical formulas [1].

B. Enter mathematical formulas

There are three ways to enter formulas in a text editor. The first is selection of elements of the formula as characters available in the selected panel. The second way is to click on

the formula editor and select characters from the context menu. The third way is to enter the formula in markup language.

- *Using the selection panel or context menu*

This method is simple, intuitive and easy to use. For example, through software OpenOffice.Org, the formula editor you can select the components to create formulas in table or through the context menu then enter the value corresponding to the formula:
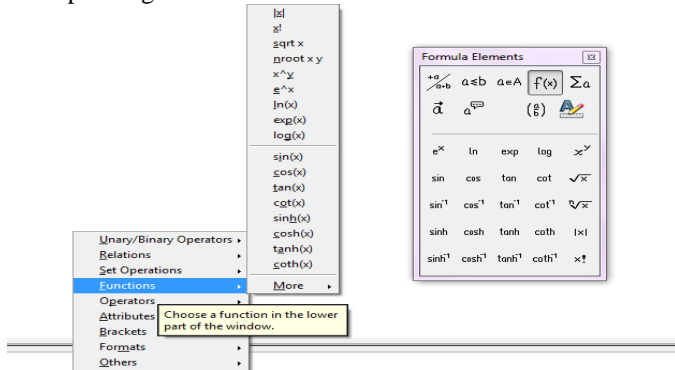

Fig. 1.   Enter the formula in the selection panel or the context menu.

- *Using markup languages*

In this way, we can enter directly the formula through markup language into the equation editor but it is not a friendly method for user. For example, to create the formula $c^2 = a^2 + b^2$ in TeXworks, we must remember the code to declare the necessary documentation and code for creating formula.
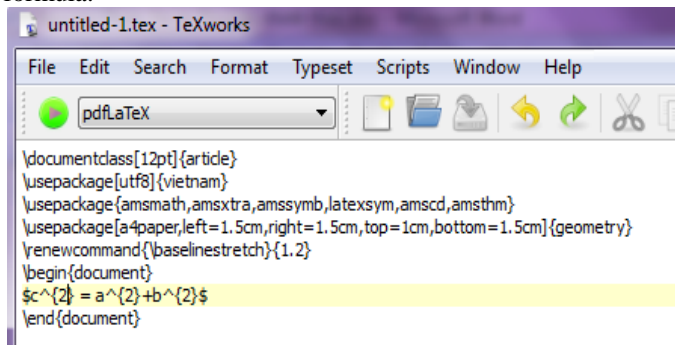

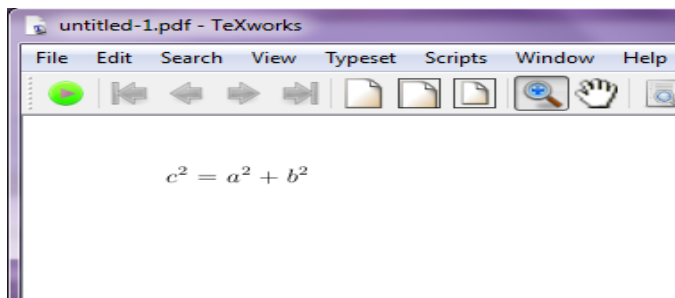Fig. 2.   Enter the formula in the format of markup language

Fig. 3.   Result represented on TeXworks

III.   DESCRIBE THE APPLICATION

From the practical requirement of the editor, we meet difficulty when search or copy mathematical formulas between editors because they use different standards. We propose development an environment editor that can solve these problems.

In this environment, we use the MathML markup language for mathematical formulas, especially in the Web environment.
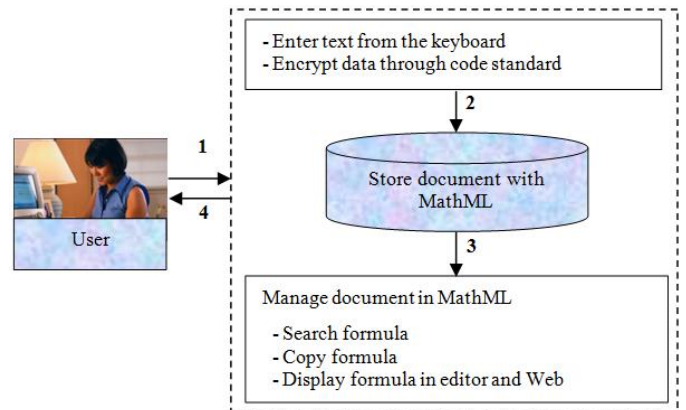

Fig. 4.   General model of the system

In this model, users can perform tasks such as drafting, editing, deleting, and searching on mathematical formulas and copy/paste directly formula from these applications into others.

This application can provide preparation, standardized data storage and perform the search, copy (especially for mathematical formulas) according to user requirements.

(1) User input from the keyboard to create mathematical formulas in text format or copy any mathematical formulas from documents created by other editing software. To do this, we develop a program to convert from markup language Methyl to other format.

(2) Storage and management of mathematical formulas as Methyl standards in writing.

(3) Provide tools to support the process of searching, copying and displaying mathematical formulas in text.

(4) Display the correct result with the requirements of the user.

IV.   PROPOSED SOLUTION

A. Content storing

To describe the process of content storing and display a given mathematical formulas. We propose solution as follows:
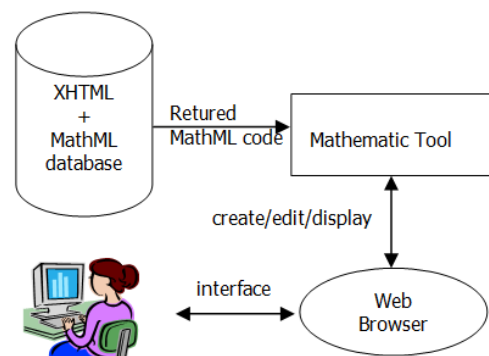

Fig. 5.   Content storing model

- Database XHTML and HTML is generated and stored in a particular host system.
- The tools of mathematical formulas were created as a combination of structured markup language HTML and MathML.
- On the Web browser, users can perform tasks to create, edit and display content from the server system.

For example, a mathematical formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ is defined in Amaya environment as the following:

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
 <mi>x</mi>
 <mo>=</mo>
 <mfrac>
  <mrow>
   <mo>&#x2212;</mo>
   <mi>b</mi>
   <mo>±</mo>
   <msqrt>
    <msup>
     <mi>b</mi>
     <mn>2</mn>
    </msup>
    <mo>&#x2212;</mo>
    <mn>4</mn>
    <msup>
     <mi>ac</mi>
    </msup>
   </msqrt>
  </mrow>
  <mrow>
   <mn>2</mn>
   <mi>a</mi>
  </mrow>
 </mfrac>
</math>
```

B. Editor environment

To edit the text, we use open source software Amaya [3]. Amaya software is WYSIWYG (What You See Is What You Get), the user can just recently drafted and can see the results displayed in the browser. The toolkit Amaya is similar Microsoft Word, OpenOffice.Org Math, …
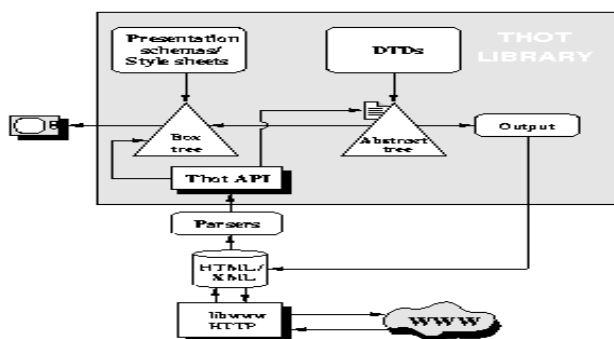

Fig. 6. The components structure of Amaya open source

- **Abstract tree**: provide information on the structural elements such as formatting headers, paragraphs,...
- **Thot API**: Thot library provides developers the API library functions, a mechanism that allows changing and expanding the function by interfering in open source.

- **DTDs**: the logical structure of the document is restricted according to certain rules specified in DTDs. These rules define how the elements are available in abstract tree can assemble to create a valid structure and the combination of elements and attributes.
- Amaya use some scheme to determine how to present different layouts for HTML documents. The schema is automatically extended with the corresponding canonical regulations CSS turned off when the special requirements involved.
- **Parsers**: Parsing the data to build the structure and logic of the abstract of the document tree for HTML, XML and CSS.
- **Lipwww HTTP**: Amaya access remote Web servers via libwww, done by HTTP / 1.1.

C. Search solution for mathematical formulas

Search engines in Maya can find special symbols in the formula but for other symbols as a sign base, integral, index, can not find directly. We add search functionality (based on the tags) and switch colors on the content found to distinguish it from other content.

We see the synchronization between sources and display the search success.
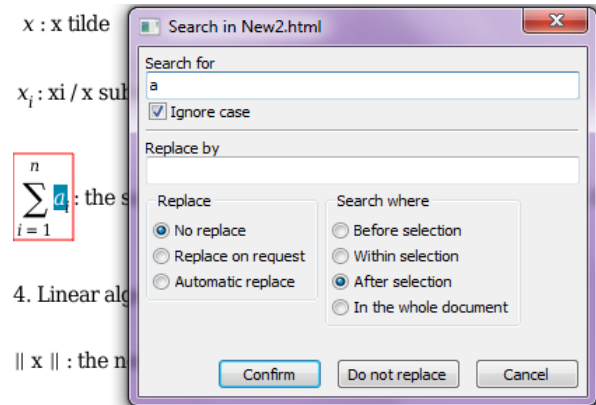

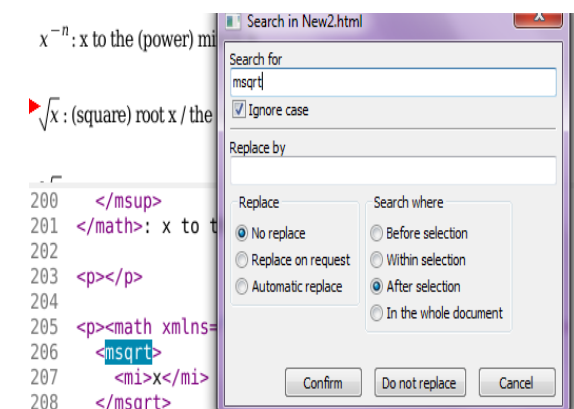Fig. 7. Searching based on character

Fig. 8. Searching based on formula

To perform a search, we have to insert the code into the source code of Amaya. Here is an example of a function void SynchronizeSourceView (NotifyElement * event) contains additional content:

```
/* record the highlighted element */
   HighlightDocument = otherDoc;
   HighlightElement = otherEl;
   HighLightAttribute = attr;
   /* if the comment is not the text that will select the element - This function
is used to change the color and type of mathematical symbols except the Text
tab */
   if (TtaGetTextLength(otherEl) <= 0)
   TtaSelectElement(otherDoc, otherEl);
   /* Scroll all views where the element appears to show it */
   for (view = 1; view < 6; view++)
   if (TtaIsViewOpen (otherDoc, view))
     {
       TtaGiveBoxAbsPosition (otherEl, otherDoc, view, UnPixel,&x, &y);
       TtaGiveWindowSize (otherDoc, view, UnPixel,&width,
       &height);
       if (y < 0 || y > height - 15)
        TtaShowElement (otherDoc, view, otherEl, 25);
     }
```

D. Solution copy mathematical formulas

Amaya is an editor and also Web browser. Therefore, all data is created to comply with the format of an XHTML page. Amaya was born with the ability to copy data tag as a string from another application into the browser but can not generate SVG tag when copying image data from other applications.

So, we built a resident program and integrated into Amaya to allow copying formulas from OpenOffice.Org Writer to Amaya browser.

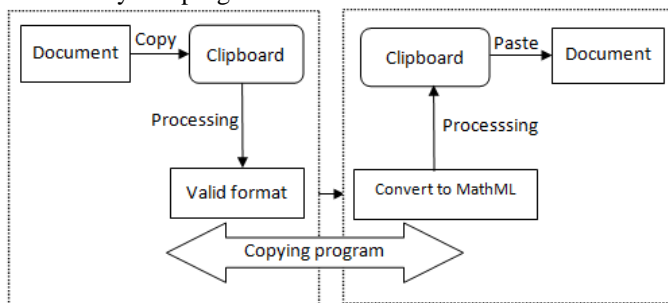The way this program works as follows:

Fig. 9. Data convert diagram in the clipboard

V. EXPERIMENT

We have developed an experimental browser based on the open source Amaya. After restarting the browser, create a new document in .html format, the screen expansion draft as follows:
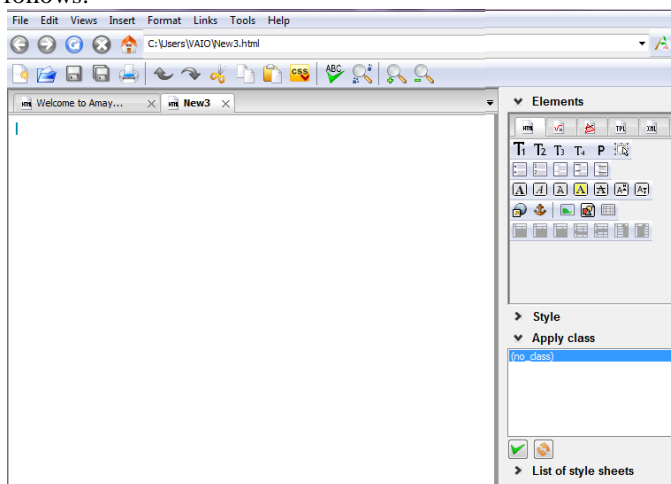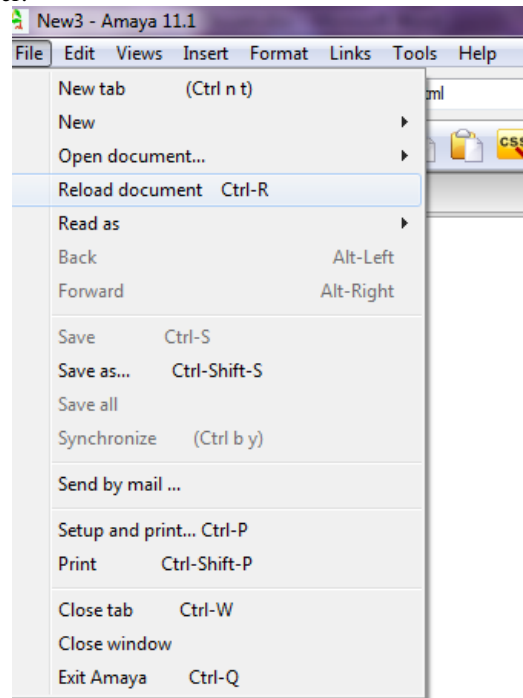

Fig. 10. Interface screen of editor

The menu bar has provided tools such as a word processing application and has provided tools such as a Web browser.

With this application, we can just edit and view the page source (code structure).

The software is also lets us search mathematical formulas. For example, the formula $x \geq y$ is presented in MathML markup language as follows:

```
<p><math xmlns="http://www.w3.org/1998/Math/MathML">
  <mrow>
    <mi>x</mi>
    <mo>&#x2265;</mo>
    <mi>y</mi>
  </mrow>
</math> </p>
```

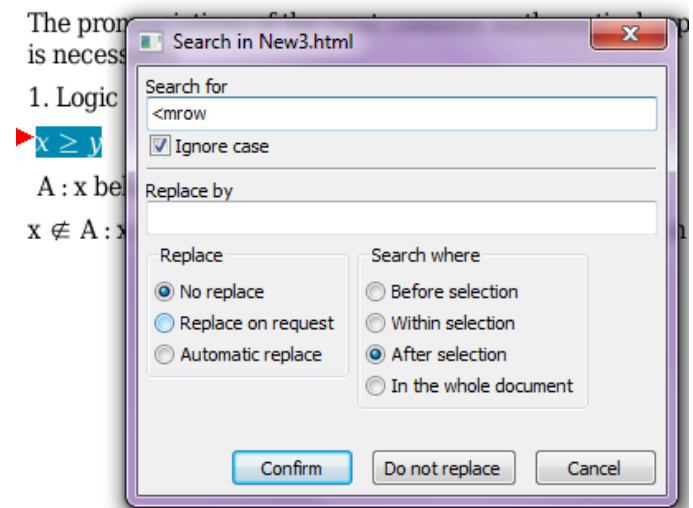The formula begins with the expression tag <mrow> and end tag </mrow>, so to find expression we just need to find the <mrow>.

Fig. 11. Search formula x ≥ y

VI. CONCLUSION

We have studied the storage standards and editing tools related to mathematical formulas. We proposed solutions to develop an environment in which the editor issues related to mathematical formulas and processed fully as possible.

The result has been that the proposed solution is used as the standard MathML to store and process mathematical formulas. We have also successfully used open source to build Amaya editor environment that supports mathematical formula editor MathML standards and develop new modules for integration into the search service Amaya formula mathematics in this environment and copy the formula from OpenOffice.org (formulas stored on other standards) into the environment our editor (using standard MathML).

However, applications in just stop at the test and used for some specific cases but untreated for all mathematical formulas. Our environment is a prerequisite for building a system editor and search for a more complete formula based on MathML and standards in the Web environment. On the basis of this study, we will continue to research more to complete the application. Further research directions, building a data warehouse storing the scientific texts and a search is stored on the server system. All these applications are hosted on the servers of the system and interact with each other based on HTTP [5].

REFERENCES

[1] D. Carlisle, P. Ion, R. Miner, Mathematical Markup Language (MathML) Version 2.0 (Second Edition), 2010

[2] D.E. Knuth, *The TeXbook,* Computers and Typesetting, Volume A, Published by Addison-Wesley, ISBN 0-201-13448-9, 1984

[3] R. Guetari, V. Quint and I. Vatton, Amaya: an Authoring Tool for the Web,
http://opera.inrialpes.fr/people/Ramzi.Guetari/Papers/Amaya.html, 2010.

[4] L. Lamport, LaTeX: A Document Preparation System, Published by Addison-Wesley, 2nd. ed., ISBN 0-201-52983-1, 1994

[5] Le Thanh Nhan, Vo Trung Hung, Cao Xuan Tuan, MATHIS – (MAThematic Information web Services) – semantic annotation and search tools for scientific resources on the web, Journal of Science and Technology, University of Danang, Volumn 4(39), 2010

[6] T. Berners-Lee, D. Connolly, HyperText Markup Language Specification, Published by IETF, 2010

[7] L. Lamport, LaTeX: A Document Preparation System, Addison-Wesley Professional, 2 edition, ISBN-13: 978-0201529838, 1994

[8] S.E. Hinton, *TEX*, Delacorte Press, ISBN-13: 978-0385375672, 2013