

BỘ DỤC VÀ ĐÀO TẠO

ĐẠI HỌC ĐÀ NẴNG

PHAN THỊ LỆ THUYỀN

**NGHIÊN CỨU ỨNG DỤNG MATHML
ĐỂ QUẢN LÝ CÁC CÔNG THỨC TOÁN
HỌC TRÊN VĂN BẢN**

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số : 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2011

MỞ ĐẦU

1. Lý do chọn đề tài

Hiện nay, có rất nhiều công cụ cho phép soạn thảo các công thức toán học trên văn bản như Winword, OpenOffice.org Writer, Acrobat, WebEditor,... Mỗi phần mềm soạn thảo văn bản có một chuẩn lưu trữ và được quản lý khác nhau. Điều này làm cho người sử dụng gặp nhiều khó khăn khi sao chép nội dung các công thức toán học giữa các ứng dụng và không thể thực hiện yêu cầu tìm kiếm một công thức toán học trên một ứng dụng soạn thảo.

Xuất phát từ nhu cầu thực tế đó, việc nghiên cứu về cách thức lưu trữ, soạn thảo, tìm kiếm các công thức toán học và đặc biệt nghiên cứu tiêu chuẩn MathML trong việc quản lý các công thức toán học trên văn bản có ý nghĩa khoa học và thực tiễn rất cao. Để có điều kiện nghiên cứu sâu hơn về MathML và ứng dụng, tôi đã chọn đề tài tốt nghiệp “Nghiên cứu ứng dụng MathML để quản lý các công thức toán học trên văn bản”.

2. Mục tiêu của đề tài

Mục tiêu chính mà đề tài hướng đến là ứng dụng ngôn ngữ đánh dấu MathML trong phần mềm mã nguồn mở để soạn thảo, quản lý, tìm kiếm công thức toán học trên văn bản và sao chép công thức toán học từ hai phần mềm soạn thảo với hai ngôn ngữ đánh dấu khác nhau. Để đáp ứng mục tiêu đã nêu, đề tài cần giải quyết những vấn đề chính sau: Nghiên cứu tổng quan về tiêu chuẩn lưu trữ, soạn thảo và hiển thị các công thức toán học trên các văn bản; Nghiên cứu một số công cụ hỗ trợ soạn thảo công thức toán học và thực trạng tìm kiếm công thức toán học hiện nay; Nghiên cứu một phần mềm mã nguồn mở để mô phỏng ứng dụng soạn

thảo công thức, lưu trữ và hỗ trợ tìm kiếm trên văn bản; Xây dựng phần mềm thực hiện chức năng sao chép công thức toán học từ hai trình soạn thảo.

3. Đối tượng nghiên cứu

Đối tượng nghiên cứu khi thực hiện đề tài là các công cụ hỗ trợ soạn thảo, các tiêu chuẩn lưu trữ, phương pháp hiển thị và tìm kiếm công thức toán học. Trong đề tài này, tôi chọn mã nguồn mở Amaya để triển khai ứng dụng.

4. Phương pháp nghiên cứu

Tìm hiểu lý thuyết về soạn thảo, lưu trữ và tìm kiếm các công thức toán học trên văn bản. Tìm hiểu chuẩn MathML để đặc tả các công thức toán học. Nghiên cứu một số công cụ hỗ trợ soạn thảo công thức toán học.

Dựa trên lý thuyết đã nghiên cứu, tiến hành xây dựng một ứng dụng soạn thảo công thức, lưu trữ, hỗ trợ tìm kiếm trên văn bản và sao chép công thức toán học từ định dạng khác.

5. Ý nghĩa khoa học và thực tiễn của đề tài

Dựa trên cơ sở lý thuyết đã nghiên cứu tiến hành xây dựng một môi trường soạn thảo, tìm kiếm và sao chép công thức toán học trên chuẩn MathML trong môi trường Web. Điều này có ý nghĩa hết sức thiết thực đối với việc trao đổi tài liệu khoa học trong thời đại công nghệ thông tin như hiện nay. Kết quả của đề tài sẽ góp phần quan trọng trong việc xử lý các công thức toán học trong các hệ thống hỗ trợ dạy và học qua mạng Internet, diễn đàn khoa học,...

6. Bố cục của luận văn

Báo cáo của luận văn được tổ chức thành 3 chương. Trong chương 1, trình bày kết quả nghiên cứu về các ứng dụng soạn thảo văn bản, các công cụ hỗ trợ soạn thảo công thức toán học và ngôn ngữ đánh dấu MathML. Chương 2 được dành để mô tả ứng dụng, xây dựng mô hình tổng quát, đề xuất giải pháp xây dựng môi trường soạn thảo công thức, tiêu chuẩn lưu trữ và tìm kiếm công thức toán học, giải pháp sao chép công thức từ OpenOffice.Org Writer sang WebBrowser Amaya. Nội dung chương 3 trình bày môi trường triển khai ứng dụng, môi trường soạn thảo, tìm kiếm công thức toán học và sao chép công thức giữa hai trình soạn thảo.

CHƯƠNG 1. NGHIÊN CỨU TỔNG QUAN

Trong thời đại ngày nay, máy tính đã trở thành công cụ không thể thiếu để phục vụ công tác soạn thảo văn bản. Hiện nay phần mềm hỗ trợ việc soạn thảo văn bản khoa học khá đa dạng, tuy nhiên các phần mềm này không đồng nhất về một chuẩn soạn thảo công thức toán học. Vì vậy, việc chuyển đổi tài liệu qua lại giữa các phần mềm rất khó khăn do đó cần phải có một chuẩn thống nhất để lưu trữ và quản lý. Trong chương này, tôi trình bày những ngôn ngữ đánh dấu để đặc tả công thức toán học trên các phần mềm soạn thảo văn bản và thực trạng tìm kiếm công thức toán học trong các tài liệu hiện nay.

1.1. SOẠN THẢO, LƯU TRỮ VÀ TÌM KIẾM CÁC CÔNG THỨC TOÁN HỌC TRÊN VĂN BẢN

1.1.1. Khái niệm văn bản

1.1.2. Hệ soạn thảo văn bản

1.1.3. Quá trình nhập công thức

1.1.3.1. Sử dụng bảng lựa chọn

Đơn giản, trực quan và dễ sử dụng.

1.1.3.2. Sử dụng ngôn ngữ đánh dấu

Viết cú pháp lệnh, đòi hỏi người sử dụng phải nhớ khi soạn thảo.

1.1.4. Tình trạng sử dụng phần mềm gõ công thức toán học hiện nay

Hiện nay, trên thế giới có rất nhiều phần mềm hỗ trợ quá trình gõ công thức toán học trên văn bản. Ngoài những phần mềm thương mại

như: Mathtype, Microsoft Equation trong bộ Microsoft Office,... hoặc những phần mềm mã nguồn mở như: MathCast, Math trong OpenOffice.Org, ... hoặc phần mềm phi thương mại của Việt Nam như MyeqText,... Nhưng mỗi công cụ chỉ hỗ trợ phù hợp cho từng phần mềm soạn thảo cụ thể. Mặt khác, với mỗi trình soạn thảo văn bản có một chuẩn lưu trữ khác nhau vì vậy sẽ gặp rất nhiều khó khăn trong việc hợp nhất các văn bản được tạo ra từ nhiều ứng dụng khác nhau.

1.1.5. Biểu diễn công thức toán học trên máy tính

1.1.5.1. Biểu diễn công thức toán học trên Microsoft Word

Ví dụ trong Microsoft word 2003, để hiển thị công thức $\sqrt[3]{x}$, soạn thảo bằng phương trình Editor thì phải sử dụng đoạn mã sau:

$\{ EQ \sqrt[3]{x} \}$

1.1.5.2. Biểu diễn công thức toán học trên WebSite

Ví dụ đoạn mã lệnh được soạn thảo trực tiếp vào cửa sổ soạn thảo của trang web tại địa chỉ <http://www.codecogs.com/latex/eqneditor.php>

Để hiển thị hình ảnh $\sqrt[3]{x}$ chúng ta phải nhập đoạn mã sau:

$\sqrt[3]{x}$

1.1.5.3. Biểu diễn công thức toán học trên OpenOffice.Org

Để hiển thị $\sqrt[3]{x}$ thì chúng ta có đoạn mã tương ứng sau:

$\sqrt[3]{x}$

1.1.6. Thực trạng tìm kiếm công thức toán học

Công thức được định dạng như một tập tin hình ảnh. Vì lý do đó, không thể thực hiện quá trình tìm kiếm như tìm kiếm chuỗi thông thường trong văn bản thường sử dụng. Các trình ứng dụng soạn thảo văn bản thông dụng hiện nay chưa có công cụ trợ giúp cho việc tìm

kiểm công thức toán học. Trên Internet, Website LatexSearch có hỗ trợ tìm kiếm các công thức toán học được soạn thảo bằng ngôn ngữ đánh dấu Latex. Đây là bản quyền của MPS Technologies (Mathematical Programming System). Nhưng những kết quả tìm thấy chỉ giới hạn trên những tài liệu điện tử trên máy chủ SpringerLink.

1.1.7. Các phần mềm hỗ trợ soạn thảo công thức toán học

1.1.7.1. Ứng dụng Microsoft Equation 3.0 trong bộ Office

Microsoft Equation là ứng dụng bộ Office hãng Microsoft bán kèm.

1.1.7.2. Phần mềm MathType

Mathtype là một phần mềm thương mại của Design Science tạo các ký hiệu toán học trên Microsoft Word, OpenOffice.Org, E-Learning, ...

1.1.7.3. Phần mềm MyEqText

MyEqText tích hợp vào Microsoft Word như một thanh công cụ.

1.2. GIỚI THIỆU MỘT SỐ NGÔN NGỮ ĐÁNH DẤU TOÁN HỌC

1.2.1. Ngôn ngữ đánh dấu LaTeX

Tex là một hệ thống sắp chữ được viết bởi Donald E.Knuth ở Đại học Stanford vào năm 1977. Nó phổ biến trong môi trường hàn lâm, đặc biệt là trong cộng đồng toán học, vật lí và khoa học máy tính.

Bắt đầu từ năm 1980, Leslie Lamport bắt đầu tạo ra hệ thống soạn thảo văn bản ngày nay gọi là LaTeX dựa trên định dạng của Tex. Hệ thống LaTeX thêm vào những lệnh tổng quát cho Tex và hướng người dùng vào cấu trúc văn bản hơn là định dạng chi tiết. Một số lệnh bậc cao cho phép người dùng dễ dàng soạn ra hầu hết các loại tài liệu. Chất

lượng bản in bằng LaTeX rất cao. Nhưng người soạn thảo không hình dung ra hình dáng của tài liệu trước khi biên dịch.

➤ Biểu diễn toán học trên văn bản

Phần nội dung toán học trong đoạn văn bản có thể được soạn thảo ở giữa dấu `\(` và `\)` hay `$` và `$` hay `\begin{math}` và `\end{math}`.

Ví dụ nội dung soạn thảo và hiển thị bằng LaTeX:

Cộng a^2 bình phương với b^2 bình phương được c^2 bình phương. Ta có thể viết dưới dạng công thức là: $c^2 = a^2 + b^2$

Cộng a bình phương với b bình phương được c bình phương. Ta có thể viết dưới dạng công thức là: $c^2 = a^2 + b^2$

1.2.2. Ngôn ngữ đánh dấu AMS-Latex

AMS-LaTeX (American Mathematical Society–LaTeX) được phát triển bởi Hội Toán học Mỹ. Tên gọi AMS-LaTeX có nghĩa là sử dụng LaTeX với phần mở rộng AMS. AMS-LaTeX ban đầu được viết bởi Michael Spivak và gần như chỉ phổ biến ở Mỹ.

ASM-LaTeX cũng là phần mềm biên dịch ra tập tin .pdf giống như LaTeX. Trong ASM-LaTeX có 2 gói là: Amsmath – gói mở rộng cho LaTeX tạo thuận lợi cho quá trình tạo công thức, nâng cao chất lượng in ấn và Amscls bao gồm lớp các tài liệu và thuật toán hỗ trợ trong quá trình soạn thảo.

➤ Biểu diễn toán học trên văn bản

Ví dụ nội dung soạn thảo và hiển thị bằng AMS-LaTeX:

```
\begin{align}
y &= (x+1)^2 \\
&= x^2 + 2x + 1
\end{align}
```

$$\begin{aligned} y &= (x + 1)^2 \\ &= x^2 + 2x + 1 \end{aligned}$$

1.2.3. Ngôn ngữ đánh dấu HTML

HTML (HyperText Markup Language - Ngôn ngữ đánh dấu siêu văn bản) là một ngôn ngữ đánh dấu được thiết kế để tạo nên các trang web. Việc đưa các công thức toán học vào tài liệu HTML mặc dù chúng ta đã có chuẩn để thực hiện điều này nhưng hầu hết các trình duyệt chưa hỗ trợ hay thiếu các phong chữ cần thiết. Do đó, chúng ta nhìn thấy công thức biểu diễn không rõ ràng, rất khó khăn trong việc thể hiện ý nghĩa toán học.

Ví dụ soạn thảo và hiển thị $\sqrt{1-e^2}$ trên Microsoft Office FrontPage 2003:

<code>&radic; 1- e<sup>2</sup></code>

$\sqrt{1-e^2}$

1.3. NGÔN NGỮ ĐÁNH DẤU MATHML

1.3.1. Lịch sử của MathML

1.3.2. Tính năng của MathML

MathML được hỗ trợ bởi các phần mềm văn phòng như Microsoft Word, OpenOffice.org cùng với các phần mềm tính toán kỹ thuật như Maple, Mathematica, và MathCad trên các nền nhiều hệ điều hành khác nhau như Linux, Windows,...

1.3.3. Cách thức hiển thị của MathML

➤ Ưu điểm

- Không chỉ hiển thị tốt mà còn chứa ý nghĩa nội dung công thức toán học.
- Dễ dàng sử dụng trên các trình duyệt.
- Cho phép hiển thị ngay công thức một cách đẹp mắt.

- Truyền tải ý nghĩa toán học cho các phần mềm tính toán.
- **Nhược điểm**
 - Cấu trúc không ngắn gọn như Tex.
 - Cần có công cụ soạn thảo thân thiện với người dùng.
- **Hiển thị trên browser**
 - Trong số các trình duyệt thân thiện như Mozilla Firefox, Amaya thì hiển thị trực tiếp MathML.
 - Các trình duyệt khác cần phần mềm hỗ trợ như: các phiên bản trước Internet Explore cài thêm MathPlayer, từ Internet Explore 6 cài JavaScript MathJax.

1.3.4. Cấu trúc của MathML

Tất cả các ký hiệu toán được đặt trong các cặp thẻ sau:

```
????<math xmlns='http://www.w3.org/1998/Math/MathML'>  
<mrow>  
    Các thẻ ký hiệu toán học  
</mrow>  
</math>  
????
```

1.3.5. Tạo công thức toán học

Tiểu kết chương 1.

Chương này đã trình bày những nghiên cứu tổng quan về biểu diễn công thức toán học trên máy tính bằng ba trình ứng dụng cụ thể. Thông qua các ví dụ, chúng ta thấy rằng mỗi phần mềm soạn thảo công thức được đặc tả bởi các ngôn ngữ đánh dấu khác nhau. Vì vậy, việc chuyển đổi tài liệu khoa học giữa các ứng dụng sẽ rất khó khăn.

Các công thức hiển thị trên màn hình soạn thảo đều ở dạng hình ảnh. Do đó, công cụ tìm kiếm không thể tìm thấy các công thức này như những chuỗi văn bản. Bên cạnh đó cũng nghiên cứu một vài ngôn ngữ đánh dấu đang được sử dụng từ đó rút ra những ưu và nhược điểm của từng ngôn ngữ.

Nhu cầu trao đổi và tìm kiếm tài liệu khoa học tự nhiên qua Internet ngày càng nhiều. Các tài liệu đăng tải lên mạng hầu hết dưới dạng tập tin và công thức hiển thị dưới dạng hình ảnh. Với việc nghiên cứu những tính năng ưu việt của ngôn ngữ đánh dấu toán học MathML trong đặc tả công thức toán học sẽ là tiền đề cho việc đề xuất giải pháp bài toán của chương 2. Đó là xây dựng một môi trường soạn thảo có hỗ trợ chuẩn MathML, mà ở đó có thể thực hiện chức năng tìm kiếm công thức toán học và xây dựng chương trình thường trú để chuyển đổi các ngôn ngữ đánh dấu khác sang ngôn ngữ đánh dấu MathML.

CHƯƠNG 2. ĐỀ XUẤT GIẢI PHÁP

Chương này sẽ đề xuất giải pháp đi xây dựng một môi trường soạn thảo văn bản trong đó có hỗ trợ chuẩn MathML đặc tả cho công thức toán học. Ở môi trường soạn thảo này người sử dụng có thể thực hiện chức năng tìm kiếm các công thức toán học trong văn bản. Và đề xuất giải pháp cho bài toán sao chép công thức toán học từ phần mềm soạn thảo văn bản này sang phần mềm soạn thảo văn bản khác và giữa hai ứng dụng trên khác nhau về chuẩn đặc tả ngôn ngữ đánh dấu toán học.

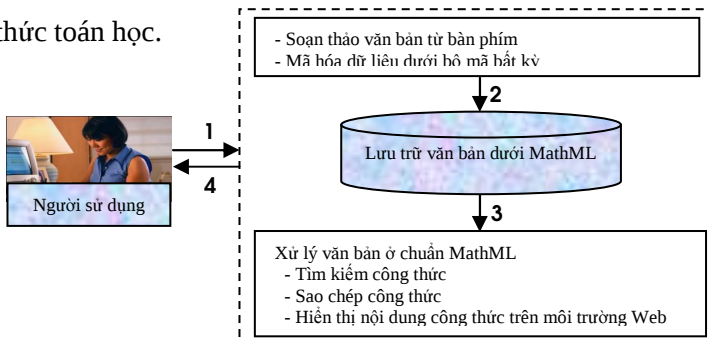
2.1. MÔ TẢ ỨNG DỤNG

Ứng dụng gồm có hai đối tượng tương tác với nhau:

- Người sử dụng : Sẽ làm nhiệm vụ soạn thảo, hiệu chỉnh, xoá, sao chép, dán văn bản và thực hiện lệnh lưu trữ dưới dạng .html do người dùng định vị, tìm kiếm công thức toán học trên văn bản.
- Môi trường soạn thảo: Sẽ cung cấp môi trường soạn thảo, chuẩn lưu trữ và thực hiện các lệnh do người sử dụng yêu cầu.

2.2. MÔ HÌNH TỔNG QUÁT

Trong môi trường soạn thảo này, tôi sử dụng MathML để đặc tả công thức toán học.



Hình 2. 1. Mô hình tổng quát của hệ thống

- (1) Người dùng nhập liệu từ bàn phím tạo các công thức toán học trên văn bản dưới dạng chuẩn MathML hoặc sao chép công thức toán học từ những tài liệu được tạo bởi phần mềm soạn thảo khác. Để làm được điều này, nhờ vào một công cụ trung gian chuyển ngôn ngữ đánh dấu đặc tả này sang ngôn ngữ đánh dấu MathML theo cấu trúc của môi trường soạn thảo.
- (2) Lưu trữ và quản lý công thức toán học dưới dạng chuẩn MathML trong văn bản.
- (3) Cung cấp một công cụ hỗ trợ cho quá trình tìm kiếm công thức toán học trong văn bản.
- (4) Chức năng tìm kiếm công thức toán học sẽ phục vụ theo yêu cầu tìm kiếm của người dùng.

2.3. ĐỀ XUẤT GIẢI PHÁP

Ngày nay, Internet là một môi trường trao đổi thông tin phổ biến. Các phương pháp dựa trên hình ảnh hiện đang chiếm ưu thế làm phương tiện truyền ký hiệu khoa học trên các trang web. Tuy nhiên, ý tưởng sử dụng hình ảnh để hiển thị công thức toán học có một số nhược điểm như sau:

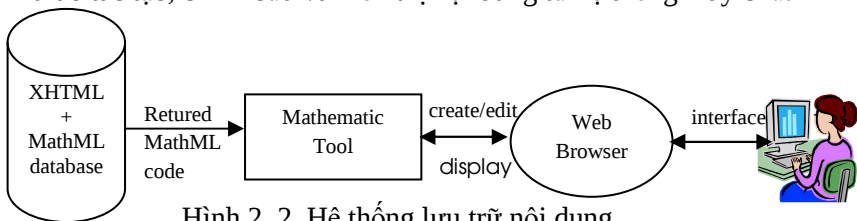
- Hình ảnh không rõ nét trong việc in ấn.
- Người sử dụng không thể sửa đổi phong chữ của công thức nên khó kế thừa trong việc sử dụng trong các ứng dụng khác.
- Tải về một trang chứa các công thức toán học tương đối chậm.
- Không thể thực hiện tìm kiếm công thức trên hình ảnh được.

Từ thực trạng trên, tôi xin đề xuất giải pháp sử dụng một chuẩn để lưu trữ nội dung, tạo một môi trường soạn thảo Web Brower và cách quản lý, tìm kiếm công thức toán học trên môi trường này.

2.3.1. Tiêu chuẩn lưu trữ nội dung

MathML là một tiêu chuẩn mở rộng của XML để biểu diễn các ký hiệu và công thức toán học, do đó có thể được nhúng bên trong XML. Để mô tả quá trình lưu trữ nội dung đến hiển thị công thức toán học tôi xin đưa ra mô hình thực hiện như sau:

- Cơ sở dữ liệu XHTML và MathML được tạo ra và lưu trữ ở một hệ thống máy chủ nào đó.
- Bộ công cụ những ký hiệu toán học được tạo ra là sự kết hợp giữa cấu trúc HTML và ngôn ngữ đánh dấu toán học MathML.
- Trên trình duyệt WebBrowser, người sử dụng có thể thực hiện các thao tác tạo, chỉnh sửa và hiển thị nội dung từ hệ thống máy chủ.



Hình 2. 2. Hệ thống lưu trữ nội dung

2.3.2. Hệ thống soạn thảo

2.3.3. Giải pháp quản lý, tìm kiếm công thức toán học

2.3.3.1. Giải pháp quản lý công thức toán học

Phần mềm Amaya quản lý nội dung dưới dạng hình cây rất rõ ràng, các cấu trúc lồng vào nhau. Amaya quản lý các thẻ của XHTML và MathML một cách có cấu trúc.

Ví dụ một công thức toán học $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ trong môi trường

Amaya được định nghĩa các tab như sau:

```
<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>a</mi>
</msup>
</msqrt>
</mrow>
<mrow>
  <mn>2</mn>
  <mi>a</mi>
</mrow>
</mfrac>
</math>
```

2.3.3.2. Giải pháp tìm kiếm công thức toán học

- Bài toán của giải pháp tìm kiếm

❖ *Bài toán so mẫu chính xác*: với việc so mẫu chính xác ta có thể đặt ra yêu cầu bài toán như sau:

Cho mẫu P độ dài m ($P = P_1P_2 \dots P_m$ – P_i là ký tự) và văn bản T độ dài n ($T = T_1T_2 \dots T_n$ – T_i là ký tự). Tìm tất cả các vị trí xuất hiện của mẫu P trong văn bản T .

❖ *Bài toán so mẫu xấp xỉ*

Tìm kiếm xấp xỉ là bài toán tìm sự xuất hiện của một mẫu trong văn bản, trong đó sự “khớp” giữa mẫu và xuất hiện của nó có thể chấp nhận

“*k lỗi*” (*k* là một giới hạn cho trước). Có thể kể ra một vài kiểu “*lỗi*”, như những lỗi đánh máy hay lỗi chính tả trong hệ thống trích rút thông tin,... Vì trong các hệ thống tin học khó có thể tránh được các “*lỗi*” nên vấn đề tìm kiếm xấp xỉ càng trở nên quan trọng.

Bài toán được phát biểu: Cho chuỗi mẫu *P* độ dài *m* và văn bản *T* độ dài *n*. Xác định độ tương tự giữa hai chuỗi *P* và *T*.

Bài toán trên chúng ta có thể đưa về tìm chuỗi con chung dài nhất (hay khúc con chung dài nhất): Một chuỗi *w* là chuỗi con hay khúc con (substring or factor) của chuỗi *T* nếu $T = uwv$ (chuỗi *u*, *v* có thể rỗng). Chuỗi *w* là khúc con chung của hai chuỗi *P*, *T* nếu *w* đồng thời là khúc con của *P* và *T*. Khúc con chung dài nhất của hai chuỗi *P* và *T*.

Input: Mẫu *P* độ dài *m*; văn bản *T* độ dài *n*

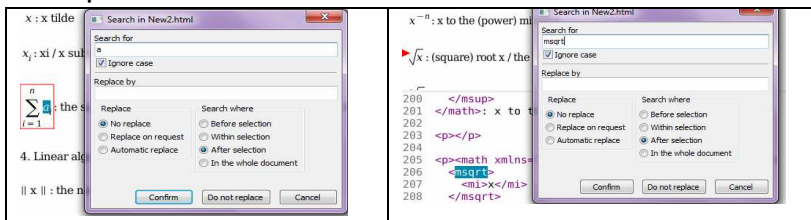
Output: - Khúc con chung dài nhất giữa *P* và *T*

- Vị trí chuỗi mẫu Substr trong *T*

- Chức năng tìm kiếm trong Amaya

❖ Chức năng tìm kiếm chuỗi trong Amaya

Amaya có thể tìm kiếm những ký hiệu đặc biệt như dấu căn, tích phân, chỉ số,... nhưng không thể tìm trực tiếp trên cửa sổ hiển thị của ứng dụng Amaya mà tìm qua mã nguồn XHTML. Và khi tìm thấy, nội dung trên cửa sổ hiển thị không được đổi màu giống như tìm kiếm thông thường mà chỉ định vị con trỏ đầu dòng của nội dung tìm thấy. Ví dụ minh họa:



Hình 2. 5. Tìm kiếm ký tự *a* và tìm kiếm ký hiệu căn bậc hai

❖ Đề xuất hướng tìm kiếm các ký hiệu toán học trong Amaya

Theo hình 2.5, ta thấy trình duyệt Amaya có sự đồng bộ giữa mã nguồn và nội dung hiển thị. Điều này có nghĩa là khi làm việc trên mã nguồn tại vị trí nào thì trên cửa sổ hiển thị sẽ được thực thi tại vị trí đó, và ngược lại. Từ phân tích trên, chứng minh rằng chức năng tìm kiếm các chuỗi trong Amaya được hỗ trợ lệnh thay đổi màu khi chuỗi được tìm thấy.

Tôi xin đề xuất bổ sung thêm chức năng chuyển đổi màu trên màn hình hiển thị khi tìm kiếm các ký hiệu toán học đặc biệt trong mã nguồn. Mã nguồn mở Amaya sau khi biên dịch có đến 11.474 tập tin, trong đó tập tin “*HTMLactions.c*” quản lý sự tương quan giữa nội dung hiển thị và mã nguồn. Trong tập tin này chứa hàm “*void SynchronizeSourceView (NotifyElement *event)*” giúp chúng ta có thể thực hiện việc chuyển đổi màu đối với các ký hiệu đặc biệt. Sau đây là một đoạn mã trong hàm *void SynchronizeSourceView (NotifyElement *event)* chứa nội dung bổ sung.

```
/* record the highlighted element */  
HighlightDocument = otherDoc;  
HighlightElement = otherEl;  
HighLightAttribute = attr;  
/* nếu những element nào không phải là text thì sẽ select  
những element đó – Chức năng này dùng đổi màu những ký  
hiệu toán học và loại trừ những thẻ là Text*/  
if (TtaGetTextLength(otherEl) <= 0)  
TtaSelectElement(otherDoc, otherEl);
```

2.3.4. Giải pháp sử dụng Clipboard sao chép công thức toán học

2.3.4.1. Làm việc với Clipboard trong OpenOffice.Org

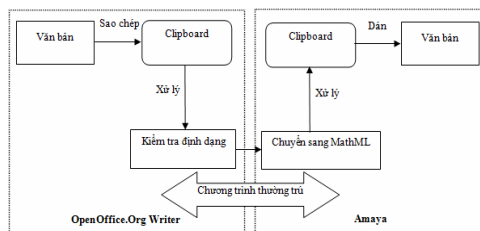
2.3.4.2. Làm việc với công thức trên Amaya

Amaya là một trình duyệt Web, vì vậy mọi dữ liệu được tạo ra phải tuân thủ theo định dạng của một trang XHTML. Amaya bắt đầu như là một trình biên tập HTML + CSS sau đó hỗ trợ XML, MathML và SVG (Scalable Vector Graphics). SVG thuộc tiêu chuẩn mở được quản lý bởi tổ chức W3C, là một ngôn ngữ đánh dấu thuộc XML dùng cho ứng dụng trên các trang mạng để miêu tả các hình ảnh đồ họa véc tơ hai chiều tĩnh và động.

Amaya có khả năng tự sinh thẻ khi sao chép dữ liệu dạng chuỗi từ ứng dụng khác vào trình duyệt nhưng không thể tự sinh ra các thẻ SVG khi sao chép dữ liệu hình ảnh.

2.3.4.3. Đề xuất giải pháp sao chép công thức từ OpenOffice.Org sang trình duyệt Amaya

Trong OpenOffice.Org Writer có hỗ trợ chức năng Save Copy As – dùng để lưu công thức thành tập tin mới dạng .mml. Tập tin này chứa các thẻ của ngôn ngữ đánh dấu MathML tương ứng với công thức vừa ghi. Nhưng cấu trúc đoạn mã không phù hợp với quy định của Amaya.



Hình 2.7. Sơ đồ quá trình chuyển đổi dữ liệu trong ClipBoard

❖ *Thuật toán chuyển đổi từ ngôn ngữ đánh dấu trong OpenOffice.Org Math thành ngôn ngữ đánh dấu MathML theo cấu trúc Amaya.*

- Ý tưởng: Thuật toán thực hiện chuyển đổi dữ liệu chuỗi từ ngôn ngữ đánh dấu OpenOffice sang ngôn ngữ đánh dấu MathML theo định dạng cấu trúc của Amaya.

+ Dữ liệu đầu vào: Ngôn ngữ đánh dấu đặc tả công thức căn bậc hai trong OpenOffice.Org.

+ Dữ liệu đầu ra: Ngôn ngữ đánh dấu đặc tả căn bậc hai theo MathML, định dạng theo cấu trúc của Amaya.

- Cài đặt thuật toán:

Theo mặc định thì khi chúng ta thực hiện thao tác sao chép, mọi dữ liệu sẽ được lưu vào Clipboard của hệ thống. Ứng dụng thường trú Math Clipboard Converter sẽ lấy dữ liệu từ Clipboard bằng lệnh “instr := Clipboard.AsText;”, sau đó xử lý theo sơ đồ thuật toán để chuyển đổi ngôn ngữ đánh dấu trong OpenOffice.Org Math sang ngôn ngữ đánh dấu MathML theo cấu trúc Amaya. Sau khi thực hiện xong việc chuyển đổi, dữ liệu cuối cùng sẽ lưu lại vào Clipboard của hệ thống bằng lệnh “Clipboard.AsText :=convert(instr);”

Tiểu kết chương 2.

Amaya là một phần mềm mã nguồn mở, miễn phí được biên dịch trong môi trường Microsoft Visual C++ 6.0 trên hệ điều hành Windows. Chương này đã đề xuất môi trường soạn thảo công thức toán học Amaya với chuẩn MathML. Trên trình duyệt Amaya, chúng ta có thể thực hiện chức năng tìm kiếm các ký hiệu toán học thông qua mã nguồn. Nhưng chức năng tìm kiếm này chỉ định vị con trỏ tại dòng văn bản trên cửa màn hình hiển thị khi tìm thấy.

Vì lý do trên, tôi đề xuất giải pháp bổ sung thêm chức năng đổi màu các ký hiệu toán học cho chức năng tìm kiếm giống như tìm kiếm chuỗi trong văn bản.

Bên cạnh đó tôi cũng đề xuất giải pháp sao chép ngôn ngữ đánh dấu đặc tả công thức toán học trong ứng dụng OpenOffice.Org Math sang ngôn ngữ đánh dấu đặc tả công thức toán học MathML theo cấu trúc của Amaya thông qua Clipboard.

CHƯƠNG 3. TRIỂN KHAI ỨNG DỤNG

Từ giải pháp của chương 2, chương này sẽ trình bày nội dung biên dịch Amaya. Kết quả thử nghiệm bổ sung đoạn mã lệnh vào mã nguồn mở Amaya với chức năng đổi màu công thức toán học khi tìm thấy và sao chép công thức toán học từ ngôn ngữ đánh dấu trong OpenOffice.Org Math sang ngôn ngữ đánh dấu MathML theo cấu trúc trình duyệt Amaya.

3.1. XÂY DỰNG MÔI TRƯỜNG SOẠN THẢO

3.2. SỬ DỤNG TRÌNH DUYỆT AMAYA

3.3. TÍCH HỢP PHẦN MỀM GỖ CÔNG THỨC TOÁN HỌC

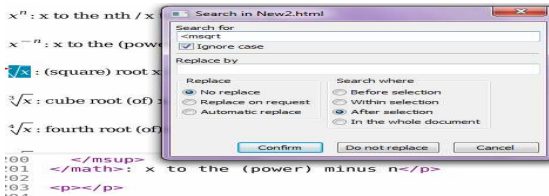
3.4. THỬ NGHIỆM

❖ *Kết quả tìm kiếm ký hiệu trong công thức toán học*

- Công thức $\sqrt{x}$, Trong Amaya được tả ngôn ngữ đánh dấu toán học MathML như sau:

```
<p><math xmlns="http://www.w3.org/1998/Math/MathML">
  <msqrt>
    <mi>x</mi>
  </msqrt>
```

Công thức được bắt đầu bằng thẻ biểu thức <msqrt> và kết thúc thẻ </msqrt>, vì vậy để tìm kiếm biểu thức trên chúng ta chỉ cần tìm theo thẻ <msqrt>. Ví dụ minh họa:

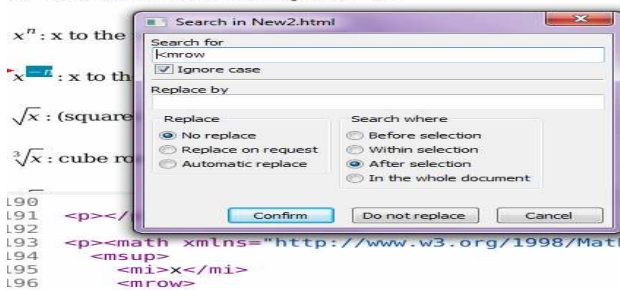


Hình 3. 8. Tìm kiếm ký hiệu toán học $\sqrt{}$

- Công thức x^{-n} , Trong Amaya được tả ngôn ngữ đánh dấu toán học MathML như sau:

```
<p><math xmlns="http://www.w3.org/1998/Math/MathML">
<msup>
  <mi>x</mi>
  <mrow>
    <mo>&#x2212;</mo>
    <mi>n</mi>
  </mrow>
</msup>
</math>
```

Công thức được bắt đầu bằng thẻ biểu thức <msup> và kết thúc thẻ </msup>, nhưng ở đây chúng ta không tìm hết công thức mà chỉ tìm chỉ số mũ $-n$. Vì vậy để tìm kiếm chỉ số trên chúng ta chỉ cần tìm theo thẻ <mrow>. Ví dụ minh họa:



Hình 3. 9. Tìm kiếm chỉ số trong toán học

- ❖ *Giao diện là một trình duyệt Web*
- ❖ *Kết quả sao chép công thức toán học*

Chương trình thường trú Math Clipboard Converter đang thông báo đã tìm thấy dữ liệu chuỗi và thực hiện chuyển đổi dữ liệu. Hình 3.13, khung bên trái là ngôn ngữ đánh dấu công thức toán học $\sqrt{5x+2}$ trong OpenOffice.Org Math và khung bên phải đoạn mã có được bằng lệnh dán từ Clipboard và hiển thị công thức tương ứng với mã lệnh.



Hình 3. 12. Kết quả của chương trình chuyển đổi Math Clipboard
Converted

Tiểu kết chương 3

Trong chương này, tôi đã trình bày nội dung kiến thức của ứng dụng Amaya và những yêu điều kiện thiết để biên dịch mã nguồn mở này. Tôi cũng đã giải quyết được những giải pháp đề xuất của chương 2 được minh họa bằng các kết quả thông qua hình ảnh thử nghiệm.

Giải pháp thứ nhất là đổi màu các ký hiệu toán học khi thực hiện chức năng tìm kiếm công thức toán học cho toàn văn bản trên môi trường Amaya.

Giải pháp thứ hai là xây dựng chương trình thường trú Math ClipBoard Converter chạy trên môi trường windows. Chương trình có chức năng chuyển đổi dữ liệu từ ngôn ngữ đánh dấu OpenOffice.Org Math sang ngôn ngữ đánh dấu MathML theo cấu trúc của Amaya. Tuy nhiên, phần mềm ứng dụng này chỉ sử dụng cho một vài trường hợp cụ thể nên còn ở mức độ trình diễn.

KẾT LUẬN

Với tốc độ phát triển mạnh mẽ của Internet, nhu cầu trao đổi tài liệu khoa học qua mạng ngày càng gia tăng mạnh mẽ. Theo thống kê của Internet World Stats khi thực hiện học tập, nghiên cứu thì con người đã chi phí khoảng 90% thời gian cho công tác tìm kiếm, phân tích và tổng hợp các tài liệu hiện có [14]. Do đó, nhu cầu tìm kiếm công thức trong các tài liệu khoa học trên môi trường Web là rất lớn. Việc nghiên cứu một công cụ soạn thảo công thức, sao chép công thức giữa các ứng dụng và tìm kiếm công thức là rất cần thiết. Điều này có ý nghĩa cao cả về mặt khoa học lẫn thực tiễn.

Về mặt lý thuyết, trong quá trình thực hiện đề tài tôi đã nghiên cứu các môi trường và công cụ soạn thảo công thức toán học đang sử dụng hiện nay. Từ đó phân tích thực trạng và khó khăn đối với người sử dụng. Nội dung chính thực hiện là tìm hiểu ngôn ngữ đánh dấu MathML, biên dịch thành công mã nguồn mở để soạn thảo công thức học bằng ngôn ngữ đánh dấu MathML, tìm kiếm công thức toán học trên môi trường này và sao chép công thức từ một ngôn ngữ đánh dấu OpenOffice.Org Math sang ngôn ngữ MathML theo định dạng của ứng dụng Amaya.

Về mặt ứng dụng, đã biên dịch thành công trình duyệt Web giúp soạn thảo công thức toán học bằng ngôn ngữ đánh dấu MathML. Bổ sung chức năng tìm kiếm vào trình duyệt và xây dựng một ứng dụng thường trú để sao chép công thức toán học từ ứng dụng OpenOffice.Org Math sang trình duyệt Amaya.

Tuy nhiên, ứng dụng trên chỉ sử dụng cho trường hợp cụ thể nên còn ở mức độ trình diễn. Trình duyệt Amaya là tiền đề cho việc xây dựng hệ thống soạn thảo và tìm kiếm công thức theo chuẩn MathML trên môi trường Web trong tương lai.

Dựa vào các kết quả này, tôi sẽ nghiên cứu thêm để hoàn thiện hơn cho các trường hợp còn lại của ứng dụng. Hướng nghiên cứu tiếp theo, là xây dựng một kho dữ liệu lưu trữ các văn bản khoa học và một bộ tìm kiếm được lưu trữ trên các máy chủ hệ thống. Tất cả các ứng dụng này đều lưu trữ trên các máy chủ của hệ thống và tương tác với nhau dựa trên nền tảng HTTP.

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG**


PHAN THỊ LỆ THUYỀN

**NGHIÊN CỨU ỨNG DỤNG MATHML
ĐỂ QUẢN LÝ CÁC CÔNG THỨC
TOÁN HỌC TRÊN VĂN BẢN**

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH

MÃ SỐ: 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

ĐÀ NẴNG - NĂM 2011

Công trình được hoàn thành tại

ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **PGS.TS. Võ Trung Hùng**

Phản biện 1: **PGS.TS. Tăng Tấn Chiến**

Phản biện 2: **TS. Nguyễn Mậu Hân**

Luận văn được bảo vệ tại Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật hợp tại Đại học Đà Nẵng ngày 10 tháng 9 năm 2011.

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng
- Trung tâm Học liệu, Đại học Đà Nẵng.