

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

**LÊ THỊ KIM ANH**

**NGHIÊN CỨU ỨNG DỤNG**  
**MÃ NGUỒN MỞ GREENSTONE ĐỂ XÂY DỰNG**  
**THƯ VIỆN SỐ TẠI TRƯỜNG ĐẠI HỌC PHÚ YÊN**

**Chuyên ngành: Khoa học máy tính**

**Mã số: 60.48.01**

**TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT**

**Đà Nẵng - Năm 2013**

Công trình được hoàn thành tại  
**ĐẠI HỌC ĐÀ NẴNG**

**Người hướng dẫn khoa học: PGS.TS. VÕ TRUNG HÙNG**

**Phản biện 1: TS. TRƯƠNG NGỌC CHÂU**

**Phản biện 2: TS. NGUYỄN MẬU HÂN**

Luận văn được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ Kỹ thuật họp tại Đại học Đà Nẵng vào ngày 18 tháng 05 năm 2013.

*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

## MỞ ĐẦU

### 1. Lý do chọn đề tài

Khi mà yêu cầu thông tin đòi hỏi nhanh chóng, chính xác, tài liệu điện tử càng thể hiện những khả năng ưu việt của mình.

Việc xây dựng thư viện số nhằm mục đích tập hợp các nguồn tài nguyên được số hoá từ những cơ sở dữ liệu trong thư viện và của trường vào một kho tài nguyên học tập tập trung. Thư viện số cung cấp kho chứa cho sự bảo tồn và duy trì các công trình khoa học, bài giảng và các bộ sưu tập điện tử, đồng thời cung cấp các công cụ phân loại cho việc truy cập, sử dụng tài nguyên và tìm kiếm tra cứu được dễ dàng tạo nên môi trường dạy, học và nghiên cứu hiệu quả.

Sau khi tìm hiểu về hệ thống những phần mềm thư viện số trong nước và trên thế giới, GreenStone nổi bật lên với tính hiệu quả, dễ sử dụng, mã nguồn mở, dễ tùy biến và mang tính chuẩn quốc tế. Rất tâm đắc với phương châm hoạt động của GreenStone và muốn xây dựng thư viện số phục vụ nhu cầu học tập và giảng dạy ở Trường, tôi chọn đề tài luận văn cao học: ***“Nghiên cứu ứng dụng mã nguồn mở GreenStone để xây dựng thư viện số tại trường Đại học Phú Yên”***

### 2. Mục tiêu nghiên cứu

Khai thác mã nguồn mở Greenstone ứng dụng trong công tác chuẩn hóa tài liệu và xây dựng các bộ sưu tập số. Đồng thời, xây dựng hệ thống thư viện số cho Trường Đại học Phú Yên với giao diện web, giúp người dùng sử dụng các bộ sưu tập cũng như các chức năng, nghiệp vụ thư viện khác.

### 3. Đối tượng và phạm vi nghiên cứu

- Đối tượng nghiên cứu: Cơ sở lý thuyết về thư viện số, các công cụ giúp xây dựng một thư viện số.

- Phạm vi nghiên cứu: mã nguồn mở GreenStone và ứng dụng xây dựng thư viện số tại Trường Đại học Phú Yên.

#### **4. Giả thiết nghiên cứu**

Hệ thống thư viện số ra đời sẽ phục vụ tốt hơn cho việc học tập và giảng dạy ở Trường Đại học Phú Yên; cung cấp các công cụ phân loại cho việc truy cập, sử dụng tài nguyên và tìm kiếm tra cứu được dễ dàng; tạo nên môi trường dạy, học và nghiên cứu hiệu quả.

#### **5. Phương pháp nghiên cứu**

- Phương pháp tài liệu: nghiên cứu các tài liệu liên quan đến thư viện số.

- Phương pháp thực nghiệm: thực nghiệm trên công cụ hỗ trợ phát triển thư viện số.

#### **6. Bố cục đề tài**

Nội dung luận văn bao gồm 3 chương được tổ chức như sau:

Chương 1: Nghiên cứu tổng quan.

Chương 2: Giải pháp xây dựng.

Chương 3: Phát triển ứng dụng

Cuối cùng là phần kết luận, hướng phát triển của luận văn.

## **CHƯƠNG 1. NGHIÊN CỨU TỔNG QUAN**

### **1.1. THƯ VIỆN VÀ THƯ VIỆN SỐ**

#### **1.1.1. Giới thiệu**

#### **1.1.2. Thư viện số**

Theo định nghĩa của Akscyn và Witten (Trường Đại học Waikato - NewZealand) thư viện số là tập hợp các bộ sưu tập số, của các đối tượng kỹ thuật bao gồm văn bản, hình ảnh, video, âm thanh cho phép:

- Truy cập, chọn lọc và hiển thị tài nguyên số (dành cho độc giả).

- Xây dựng, tổ chức và lưu hành (dành cho cán bộ thư viện).

## **1.2. THƯ VIỆN SỐ GREENSTONE**

### **1.2.1. Giới thiệu**

Đứng trước yêu cầu thực tế, năm 1995, một nhóm giảng viên và sinh viên trường Đại học Waikato – NewZealand đã xây dựng phần mềm thư viện số GreenStone. Thấy được nghĩa và tác dụng, tháng 8 năm 2000, UNESCO và Human Info NGO đã tham gia hỗ trợ và phát triển GreenStone. GreenStone là bộ phần mềm giúp người sử dụng dễ dàng xây dựng và phân phối bộ sưu tập thư viện số, nó cung cấp phương pháp mới để tổ chức thông tin và xuất bản thông tin trên Internet và qua CD ROM. GreenStone là phần mềm mã nguồn mở mang tính quốc tế được cung cấp trên <http://www.greenstone.org> với mục đích cung cấp cho các trường Đại học, thư viện và các viện nghiên cứu xây dựng các bộ sưu tập cho riêng mình.

### **1.2.2. Các khái niệm cơ bản trong GreenStone**

#### ***a. Tài liệu***

GreenStone hỗ trợ các loại tài liệu dạng HTML, XML, TXT và các dạng phức tạp như Word, RTF hoặc dạng đang được sử dụng phổ biến trên nhiều môi trường như PDF, PostScript, dạng multi-media như âm thanh (ví dụ .mp3), hình ảnh, phim...

#### ***b. Bộ sưu tập***

Một thư viện số do GreenStone tạo ra chứa được nhiều bộ sưu tập. Mỗi bộ sưu tập tập trung vào một chủ đề nào đó. Ví dụ, bộ sưu tập Sách, bộ sưu tập Luận văn... Các bộ sưu tập có thể được bổ sung cập nhật, kích thước các bộ sưu tập có thể lên đến hàng Gigabyte dữ liệu. Bộ sưu tập có thể xem là đơn vị của một thư viện số GreenStone.

#### ***c. Tìm kiếm***

Các bộ sưu tập cho phép tìm kiếm trên toàn bộ nội dung văn bản hoặc có thể tìm kiếm trên từng vùng (section) hay đoạn (paragraph). Cũng có thể tìm kiếm theo các từ khóa, hay các cụm từ và kết quả sẽ được sắp xếp theo yêu cầu của câu truy vấn.

#### ***d. Duyệt tài liệu***

GreenStone cho phép định nghĩa trước các cấu trúc để duyệt tài liệu trong mỗi bộ sưu tập dựa trên những metadata tìm thấy trong bộ sưu tập đó.

#### ***e. MetaData***

Là thông tin mô tả cho một tài liệu trong bộ sưu tập, ví dụ tên tài liệu, tên tác giả, ngày xuất bản...

GreenStone dùng các thẻ XML để mô tả thông tin cho tài liệu, ví dụ:

```
<Metadata name="Title">Tìm hiểu phần mềm GreenStone</Metadata>
<Metadata name="Author">Kim Anh</Metadata>
```

Các thẻ này có thể:

- Được nhúng trong tài liệu của bộ sưu tập (ví dụ như các thẻ HTML trong tài liệu HTML).
- Được lưu thành tập tin Metadata kèm theo tài liệu.
- Được trích một cách tự động từ một tài liệu nào đó, ví dụ thông tin về tên, kích thước, ngày tạo, ngày hiệu chỉnh,... tập tin tài liệu.

#### ***f. Biên mục***

Biên mục là khái niệm của nghiệp vụ thư viện để chỉ hành động cung cấp thông tin mô tả cho các tài liệu trong thư viện. Hiện nay người ta thường biên mục tài liệu theo chuẩn quốc tế Dublin Core.

#### ***g. Plugin***

Plugin là một chương trình con (script) được dùng trong quá trình xây dựng bộ sưu tập.

Do nguồn vào có nhiều dạng tài liệu khác nhau (PDF, Word, Text, ...) nên cần plugin để chuyển chúng về một loại thống nhất là XML của GreenStone và trích thông tin từ tài liệu nguồn đưa vào tập tin XML này. Ví dụ tài liệu nguồn là tập tin Word thì ta dùng Plugin WordPlug.

Các Plugin được viết bằng ngôn ngữ Perl. Mọi plugin đều kế thừa từ plugin cơ sở BasPlug. Plugin cơ sở BasPlug thực hiện những thao tác cơ bản như tạo tài liệu mới XML theo định dạng của Greenstone, gán định danh cho tài liệu. Các plugin được đặt trong thư mục “greenstone\perl\plugins”.

#### ❖ Các Plugin xử lý tài liệu độc quyền

Đối với tài liệu độc quyền như Word, PDF, ta dùng các plugin tương ứng là WordPlug và PDFPlug. Các plugin này thực hiện 2 thao tác:

1. Chuyển tài liệu nguồn sang dạng HTML hay Plain Text
2. Sử dụng plugin HTMLPlug hay TEXTPlug chuyển kết quả ở bước 1 sang dạng XML của Greenstone.

Để chuyển tài liệu nguồn sang dạng HTML hay Plain Text, Greenstone dùng những chương trình có sẵn như **pdftohtml**, **wwware** trong thư mục “greenstone\bin\windows”.

#### **Danh sách các Plugin**

**Bảng 1.1. Danh sách các Plugin**

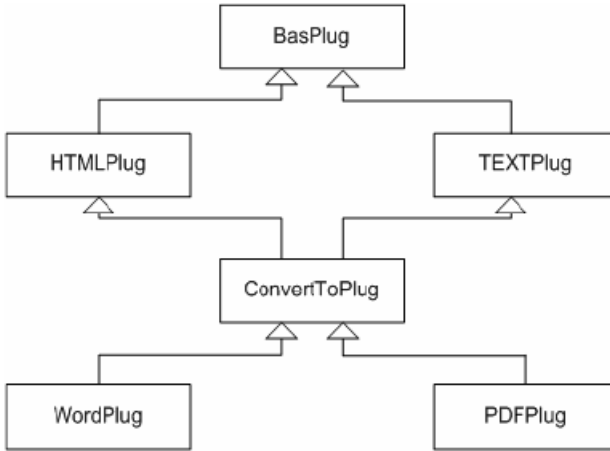
Tên Plugin	Công dụng	Kiểu tập tin xử lý	Các tập tin không xử lý
BasPlug	Là lớp cơ sở cho tất cả các plugin		
ConvertToPlug	Gọi các chương trình để chuyển các tài liệu độc quyền sang HTML hay Plain Text		
ArcPlug	Xử lý các tập tin chỉ ra trong tập archives.inf, tập tin archives.inf là cầu nối giữa tiến trình import và tiến trình build. Plugin này bắt buộc phải khai báo trong tập tin cấu hình.		
RecPlug	Duyệt qua thư mục để xử lý các tập tin mà plugin này tìm thấy.		
GAPlug	Xử lý các tập tin của Greenstone được phát sinh từ chương trình import.pl	.xml	
TextPlug	Xử lý tập tin Text thuần túy.	.txt, .text	



HTMLPlug	Xử lý tập tin HTML	.htm, .html, .cgi, .php, .asp, .shm, .shtml	.gif, .jpg, .jpeg, .png, .css, .rtf
WordPlug	Xử lý tập tin Word	.doc	.gif, .jpg, .jpeg, .png, .css, .rtf
PDFPlug	Xử lý tập tin PDF	.pdf	.gif, .jpg, .jpeg, .png, .css, .rtf
PSPlug	Xử lý tài liệu postscript, trích thông tin metadata ngày, tựa đề, số trang, ...	.ps	.eps
EMAILPlug	Xử lý thông điệp email, trích thông tin như tác giả, ngày, chủ đề, ...	Tập tin kết thúc bằng số hoặc số theo sau là .Email	
BibTexPlug	Xử lý các tập tin bibliography theo chuẩn Bib Tex	.bib	
ReferPlug	Xử lý các tập tin bibliography theo chuẩn Refer	.bib	

SRCPlug	Xử lý các tập tin mã nguồn	Makefile, Readme, .c, .cc, .cpp, .h, .hpp, .pl, .pm, .sh	.o, .obj, .a, .so, .dll
ImagePlug	Xử lý các tập tin ảnh. Plugin này chỉ dùng trên Unix	.gif, .jpg, .jpeg, .png, .bmp, .xbm, .tif, .tiff	
SplitPlug	Giống BasPlug và ConvertToPlug. Không dùng trực tiếp plugin này, plugin này phải được thừa kế để xử lý tài liệu.		
FoxPlug	Xử lý các tập tin FoxBase	.dbt, .dbf	
ZipPlug	Xử lý các tập tin nén	.gzip, .bzip, .tar, .zip, .gz, .bz, .tgz, .taz	

Các plugin WordPlug và PDFPlug kế thừa từ plugin ConvertToPlug. Tùy chọn **convert\_to** của plugin ConvertToPlug cho biết chuyển sang dạng tài liệu nào.



**Hình 1.3. Cây kế thừa của các plugin xử lý các tài liệu đọc quyền**

#### *h. Classifier*

Classifier dùng để xây dựng cấu trúc duyệt tài liệu trên web của một bộ sưu tập. Tương tự các plugin, các classifier được đặc tả trong tập tin cấu hình `collect.cfg` của mỗi bộ sưu tập.

Trong pha cuối cùng của quá trình xây dựng bộ sưu tập (nén và tạo chỉ mục trên tài liệu), các classifier được script `buildcol.pl` gọi sẽ lưu cấu trúc duyệt tài liệu vào cơ sở dữ liệu bộ sưu tập.

#### *i. Định dạng cách hiển thị tài liệu*

Những trang web trong Greenstone không được thiết kế trước mà được phát sinh và hiển thị ra web browser. Một phần giao diện của các trang web này được quản lý bởi các chuỗi định dạng. Chuỗi định dạng được đặc tả trong tập tin cấu hình của bộ sưu tập `collect.cfg`. Để đặc tả chuỗi định dạng ta dùng từ khóa **format**, theo sau là tên của những thành phần mà chuỗi định dạng sẽ tác động.

Ta có thể định dạng 2 thành phần sau:

- Danh sách tài liệu được phát sinh bởi classifier hoặc danh sách tài liệu nhận được trong quá trình tìm kiếm.
- Những thành phần trên trang web hiển thị tài liệu hoặc hiển thị các đoạn của một tài liệu.

### **1.2.3. Tính năng**

Những đặc trưng nổi bật và ưu thế của GreenStone:

- Truy cập qua trình duyệt web, cả ở chế độ cục bộ (local) và từ xa (remote).
- Chạy được trên nhiều hệ điều hành: Windows, Unix, Macintosh.
- Tìm kiếm toàn văn bản và tìm kiếm theo từng trường riêng biệt.
- Khả năng trình duyệt linh động, đa dạng.
- Cấu trúc duyệt tài liệu được xây dựng hoàn toàn tự động.
- Tận dụng các metadata sẵn có trong tài liệu, giúp người tạo lập bộ sưu tập không phải làm bằng tay.
- Khả năng linh động, dễ mở rộng hệ thống nhờ các thành phần như plugin, classifier.
- Hỗ trợ xử lý tài liệu với nhiều loại ngôn ngữ.
- Cung cấp giao diện đa ngôn ngữ.
- Ngoài các bộ sưu tập văn bản, hình ảnh thông thường, GreenStone còn cho phép tạo các bộ sưu tập hình ảnh, âm thanh đa phương tiện (multimedia).
- Khả năng lưu trữ rất lớn, tới hàng Gigabyte dữ liệu.
- Thêm mới bộ sưu tập đơn giản, có hiệu quả tức thì.
- Khả năng xuất bản các bộ sưu tập ra CD, với đầy đủ tính năng để có thể tự cài đặt và chạy độc lập.

- Các bộ sưu tập dễ dàng được mang chuyển, phân phối, chia sẻ.
- Theo phong cách WYSIWYG (What you see is what you get), dễ dùng, thuận tiện.

### **1.3. MỘT SỐ THƯ VIỆN SỐ ĐÃ CÓ**

#### **1.3.1. Thư viện iLib của CMC**

#### **1.3.2. Thư viện VeBrary của Lạc Việt**

#### **1.3.3. Thư viện Libol của Tinh Vân**

## **CHƯƠNG 2. GIẢI PHÁP XÂY DỰNG**

### **2.1. MÔ TẢ ỨNG DỤNG**

#### **2.1.1. Giới thiệu về công tác thư viện tại Đại học Phú Yên**

#### **2.1.2. Đối tượng sử dụng hệ thống**

Giảng viên, sinh viên, thủ thư.

- Đối với người dùng (giảng viên, sinh viên) truy cập vào hệ thống Web để xem thông tin điều lệ của thư viện; thực hiện tra cứu các bộ sưu tập tài liệu; xem thông báo từ thư viện; đăng nhập, đăng xuất, thay đổi mật khẩu; xem các website thư viện số khác.
- Đối với cán bộ thủ thư: sử dụng hệ thống để biên mục tài liệu và xây dựng các bộ sưu tập số.

#### **2.1.3. Các chức năng chính**

##### ***a. Xây dựng bộ sưu tập***

Thủ thư có quyền tạo các bộ sưu tập, quản lý và xuất bản bộ sưu tập lên hệ thống

##### **❖ Xây dựng bộ sưu tập Luận Văn**

Thủ thư lấy thông tin luận văn và các tập tin luận văn do các khoa chuyên môn trong trường cung cấp để tạo bộ sưu tập luận văn.

##### **❖ Xây dựng bộ sưu tập Sách**

Thông tin về Sách, Giáo trình từ cơ sở dữ liệu, thủ thư dùng để biên mục và tạo các bộ sưu tập Sách, Giáo trình.

### ❖ **Xây dựng bộ sưu tập Tổng quát**

Nhằm đáp ứng tính mở rộng của thư viện, hệ thống cho phép thủ thư tạo bộ sưu tập tổng quát. Thủ thư biên mục tài liệu theo các chuẩn biên mục, phổ biến là chuẩn Dublin Core, hoặc tự tạo ra các chuẩn riêng nhờ hiệu chỉnh các bộ metadata. Các tài liệu bao gồm: các tài liệu văn phòng (Word, Power Point, Excel...), văn bản, Acrobat Reader (PDF), HTML, XML, Email, tập tin nén (.ZIP)... hoặc âm thanh (mp3), hình ảnh, video... Sau khi biên mục, thủ thư có thể xây dựng bộ sưu tập.

#### ***b. Tra cứu tài liệu***

Người dùng thông qua website có thể tra cứu, tìm kiếm tài liệu trong các bộ sưu tập số.

## **2.2. MÔ HÌNH ĐỀ XUẤT**

### **2.2.1. Giới thiệu**

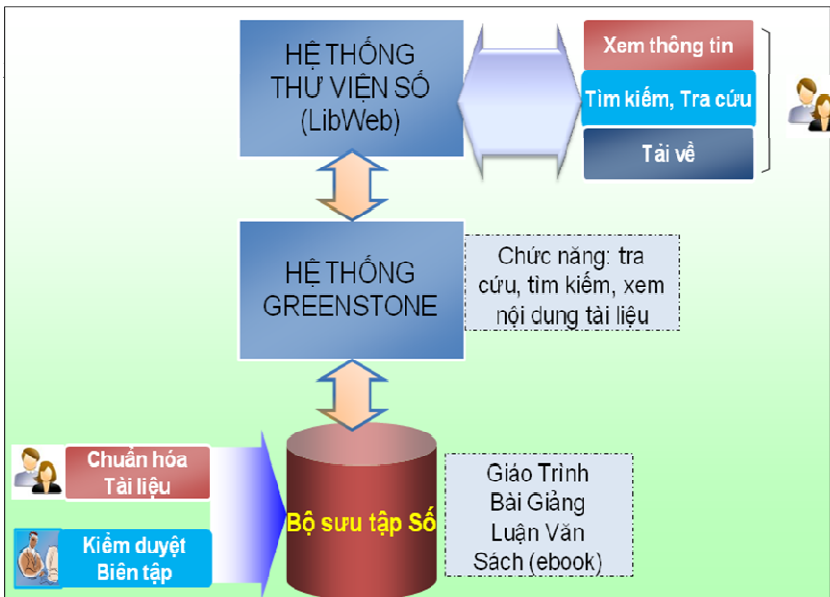
Hệ thống cho phép người dùng xây dựng bộ sưu tập luận văn, sách, giáo trình, các tài liệu khác; tìm kiếm thông tin tài liệu, tra cứu trên nội dung tài liệu, quản lý người dùng và một số chức năng khác.

#### ❖ **Người dùng (giảng viên, cán bộ, sinh viên)**

- Xem thông tin điều lệ của thư viện: nội quy, nguyên tắc làm thẻ đọc giả, các hình thức phạt
- Tra cứu các bộ sưu tập tài liệu. Ở chức năng này hệ thống LibWeb sẽ kết nối với hệ thống GreenStone để thực hiện các chức năng tra cứu, tìm kiếm tài liệu, xem nội dung tài liệu.

- Xem thông báo từ thư viện bao gồm: thông báo tài liệu mới, thông báo độc giả mượn quá hạn.
  - Thực hiện các chức năng đăng nhập, đăng xuất, thay đổi mật khẩu.
  - Xem các website báo chí, âm nhạc, hội họa.
  - Xem các website thư viện điện tử khác bao gồm thư viện điện tử của các Trường Đại học, thư viện điện tử của các tổ chức cơ quan trong và ngoài nước.
- ❖ **Cán bộ thủ thư:** sử dụng phần mềm GreenStone xây dựng bộ sưu tập số. Bộ sưu tập này được cấu hình và định dạng để

### 2.2.2. Mô hình



**Hình 2.1. Mô hình hệ thống quản lý thư viện**

Hệ thống thư viện số LibWeb là hệ thống thư viện số Trường Đại học Phú Yên được thiết kế trên môi trường Web, có liên kết với hệ thống GreenStone. Hệ thống cho phép:

- Người dùng (nói chung): Xem thông tin; tra cứu các bộ sưu tập tài liệu, ở chức năng này hệ thống LibWeb sẽ kết nối với hệ thống GreenStone để thực hiện các chức năng tra cứu, tìm kiếm tài liệu, xem nội dung tài liệu.
- Thủ thư: xây dựng bộ sưu tập cho thư viện số.

## **2.3. PHÂN TÍCH HỆ THỐNG THƯ VIỆN SỐ**

### **2.3.1. Mô hình Use Case**

### **2.3.2. Đặc tả các Use case chính**

### **2.3.3. Phân tích các Use case chính**

## **2.4. GIẢI PHÁP XÂY DỰNG BỘ SƯU TẬP**

### **2.4.1. Giới thiệu**

Quá trình xây dựng một bộ sưu tập trải qua 3 pha chính:

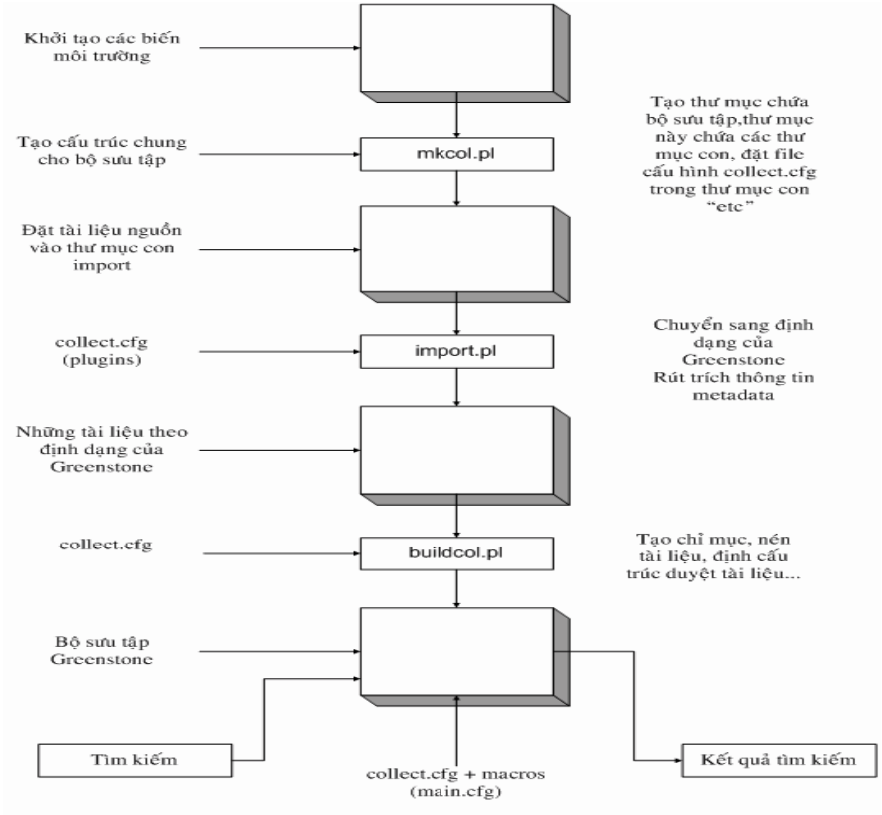
- Pha 1: Tạo cấu trúc chung cho bộ sưu tập
- Pha 2: Chuyển định dạng tài liệu nguồn sang định dạng XML
- Pha 3: Nén và tạo chỉ mục trên các tài liệu của bộ sưu tập

Ở mỗi pha ta dùng chương trình do Greenstone hỗ trợ để xây dựng bộ sưu tập :

- Pha 1: dùng chương trình `mkcol.pl`
- Pha 2: dùng chương trình `import.pl`
- Pha 3: dùng chương trình `buildcol.pl`

Các chương trình trên được đặt trong thư mục “**greenstone\bin\script**”. Để thực thi các chương trình này trong môi trường DOS, ta dùng cú pháp lệnh như sau: `perl -S <tên chương trình><các tham số>`

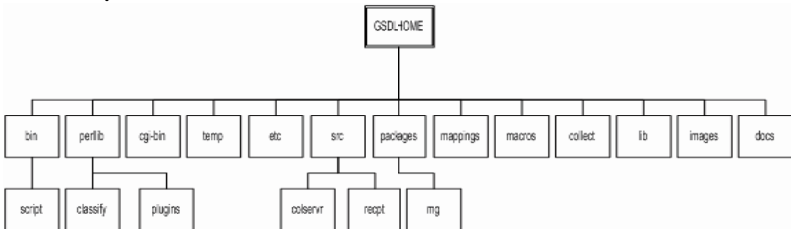




**Hình 2.10. Quá trình xây dựng bộ sưu tập**

**2.4.2. Cấu trúc thư mục của GreenStone**

Ta gọi thư mục cài đặt Greenstone là GSDLHOME. Cấu trúc thư mục của Greenstone như sau:



**Hình 2.11. Cấu trúc thư mục của Greenstone**

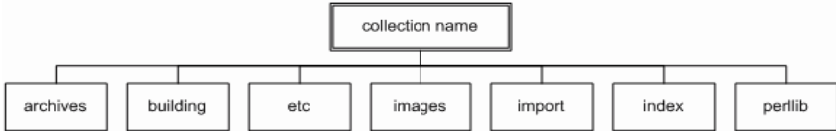
**Bảng 2.3. Danh sách thư mục của GreenStone**

Tên thư mục	Mục đích
bin	Chứa mã thực thi
bin\script	Chứa các script Perl để tạo bộ sưu tập
perllib	Chứa những module Perl hỗ trợ cho quá trình xây dựng bộ sưu tập
Perllib\plugin	Chứa mã nguồn của các plugin hỗ trợ xử lý tài liệu
Perllib\classify	Chứa mã nguồn các classify hỗ trợ việc hiển thị kết quả tìm kiếm tài liệu
cgi-bin	Chứa các CGI script của GreenStone
tmp	Chứa các tập tin tạm của GreenStone
etc	Chứa các tập tin cấu hình, tập tin log, cơ sở dữ liệu quản lý người dùng
src	Chứa mã nguồn C++
src/colservr	Chứa mã nguồn C++
src/recpt	Chứa mã nguồn C++
packages	Chứa mã nguồn của những gói phần mềm hỗ trợ cho GreenStone
Packages\mg	Chứa mã nguồn của MG - phần mềm dùng để nén và tạo chỉ mục trong GreeStone
mappings	Chứa bảng chuyển đổi chuẩn Unicode thành các chuẩn khác
macros	Chứa các tập tin macro dùng cho giao diện GreenStone
collect	Chứa các bộ sưu tập
lib	Chứa mã nguồn C++ dùng cho collection server và receptionist

images	Chứa các tập tin ảnh dùng cho giao diện của GreenStone
docs	Chứa các tài liệu về GreenStone

### 2.4.3. Cấu trúc thư mục của một bộ sưu tập

Trong pha 1 của quá trình xây dựng bộ sưu tập, sau khi thực thi chương trình mkcol.pl, bộ sưu tập được tạo ra với cấu trúc các thư mục như sau:



**Bảng 2.4. Danh sách thư mục của bộ sưu tập**

Tên thư mục	Mục đích
archives	Chứa các tập tin sau khi import
building	Chứa các tập tin trong quá trình nén, tạo chỉ mục, cơ sở dữ liệu cho bộ sưu tập
etc	Chứa tập tin cấu hình collect.cfg
images	Chứa các ảnh dành riêng cho bộ sưu tập
import	Chứa các tài liệu nguồn cần xây dựng bộ sưu tập
index	Chứa các tập tin sau khi nén, tạo chỉ mục, cơ sở dữ liệu lấy từ thư mục building
perllib	Chứa các thư viện Perl hỗ trợ cho bộ sưu tập.

### 2.4.4. Cấu trúc tài liệu theo định dạng XML

Trong pha import, Greenstone chuyển tài liệu nguồn sang tài liệu XML.

Tài liệu XML của Greenstone có một thẻ gốc là `<Section> </Section>`. Tài liệu được chia thành nhiều vùng (section) được bao bọc bởi cặp thẻ `<Section> </Section>`,

các cặp thẻ `<Section>` `</Section>` có thể lồng nhau. Mỗi Section có một thẻ Description và một thẻ Content. Thẻ Description có thể chứa một hay nhiều thẻ Metadata. Các thẻ metadata có cấu trúc:

```
<Metadata name = "Tên metadata">Giá trị của metadata</metadata>
```

Ta thường biên mục tài liệu theo chuẩn Dublin Core, ví dụ:

```
<Metadata name = "dc.Title">Tìm hiểu nguồn mở Greenstone</Metadata>
```

dc là từ viết tắt của cụm từ Dublin Core

Greenstone thiết kế sẵn một số bộ thẻ để biên mục, với Dublin Core ta có bộ thẻ dc. Nếu không có metadata nào trong các chuẩn của Greenstone thích hợp để mô tả tài liệu, có thể dùng các bộ metadata do ta định nghĩa.

Cấu trúc phân cấp của tài liệu được dùng cho chỉ mục tìm kiếm tài liệu, có 3 mức chỉ mục: `document`, `section`, `paragraph`.

Chỉ mục `document`: tìm kiếm một số từ trong tất cả các tài liệu.

Chỉ mục `section`: tìm kiếm một số từ trong từng section.

Chỉ mục `paragraph`: xem mỗi đoạn văn như là một tài liệu riêng biệt, thích hợp cho mục đích tìm kiếm tập trung.

#### **2.4.5. Tập tin cấu hình bộ sưu tập**

Tập tin cấu hình của mỗi bộ sưu tập có tên `collect.cfg` được đặt trong thư mục "**greenstone\collect<tên bộ sưu tập>etc**" dùng để quản lý giao diện, cách thức xử lý tài liệu, cách hiển thị nội dung tài liệu...

Trong quá trình xây dựng bộ sưu tập, khi ta thực thi chương trình `mkcol.pl`, một tập tin cấu hình đơn giản cho bộ sưu tập được tạo, chứa các giá trị mặc định cho bộ sưu tập.

Greenstone còn hỗ trợ chức năng tìm kiếm tài liệu trên nhiều bộ sưu tập, định nghĩa như sau:

**supercollection** <tên bộ sưu tập 1> <tên bộ sưu tập 2> ...<bộ sưu tập n>

Khi này, trong quá trình tìm kiếm tài liệu, Greenstone sẽ tìm trong cả  $n+1$  bộ sưu tập: bộ sưu tập hiện tại, bộ sưu tập 1, bộ sưu tập 2, ..., bộ sưu tập n

## 2.5. HỆ THỐNG WEB GREENSTONE

### 2.5.1. Thiết kế giao diện

Để thiết kế giao diện ta thao tác trên các tập tin sau:

- Tập tin cấu hình của một bộ sưu tập `collect.cfg`
- Những tập tin macro: có phần mở rộng là `.dm`, lưu trong thư mục "**greenstone\macros**". Macro là ngôn ngữ dành riêng cho Greenstone để xử lý giao diện web. Một macro có định dạng sau: `_tên_macro_`. Ví dụ: `_imagecollect_` là một macro.

Các trang web của Greenstone không được thiết kế sẵn, các trang này được phát sinh từ các tập tin macro. Ví dụ tập tin `home.dm` sẽ phát sinh ra trang chủ của Greenstone, tập tin `help.dm` sẽ phát sinh ra trang trợ giúp của Greenstone...

Mỗi tập tin macro chứa một hay nhiều package. Mỗi package chứa một loạt các macro. Một macro có thể được viết cho một bộ sưu tập với từ khóa [`c=tên bộ sưu tập`], nghĩa là ta sẽ viết đề macro mặc định của Greenstone.

- Những tập tin ảnh: Tất cả những tập tin ảnh được chứa trong thư mục “greenstone\images”.
- Tập tin `main.cfg`: Chứa các khai báo các tập tin macro được dùng trong Greenstone. Nếu tạo tập tin macro mới, để tập tin macro này có hiệu lực, cần thêm tên tập tin macro đó vào nội dung tập tin `main.cfg`. Tập tin `main.cfg` được lưu trong thư mục “greenstone\etc”.

### 2.5.2. Cài đặt và cấu hình lại

- a. Ảnh tiêu đề bộ sưu tập
- b. Các nút duyệt trang
- c. Ảnh tiêu đề trang web
- d. Các nút duyệt tài liệu
- e. Hiển thị văn bản
- f. Viết đề các Macro
- g. Thêm một trang mới
- h. Hiển thị các bộ sưu tập
- i. Macro chuẩn
- j. Lưu ý

## CHƯƠNG 3. PHÁT TRIỂN ỨNG DỤNG

### 3.1. QUI TRÌNH PHÁT TRIỂN ỨNG DỤNG

- Bước 1: Xây dựng các bộ sưu tập số.
- Bước 2: Thiết kế giao diện Web của GreenStone phục vụ tra cứu, tìm kiếm và xem thông tin tài liệu trong các bộ sưu tập số.
- Bước 3: Thiết kế các trang chức năng của hệ thống LibWeb.

### 3.2. XÂY DỰNG BỘ SƯU TẬP SỐ

Phần này trình bày các bước sử dụng phần mềm GreenStone để tạo ra các bộ sưu tập số (luận văn tốt nghiệp, sách/giáo trình, bài

giảng). Các bộ sưu tập này được biên mục, cấu hình và định dạng để thuận tiện cho việc tra cứu.

### 3.2.1. Xây dựng một bộ sưu tập với GLI

#### *a. Giao diện thủ thư GreenStone Librarian Interface (GLI)*

Giao diện Greenstone Library Interface [11] cung cấp giao diện tương tác phản ánh các bước thực hiện như sau:

- **GATHER:** Dùng để thu gom tài liệu tập trung vào bộ sưu tập.
- **ENRICH:** Trình bày 15 yếu tố của Dublin Core để biên mục tài liệu. Động tác biên mục được làm thủ công. Trong đó, 15 yếu tố của chuẩn Dublin Core:
  1. Nhan đề (Title): Nhan đề của tài liệu.
  2. Tác giả (Creator): Tác giả của tài liệu, bao gồm cả tác giả cá nhân và tập thể.
  3. Chủ đề (Subject): Chủ đề tài liệu đề cập dùng để phân loại tài liệu. Có thể thể hiện bằng từ, cụm từ/(Khung chủ đề), hoặc chỉ số phân loại/ (Khung phân loại).
  4. Tóm tắt (Description): Tóm tắt, mô tả nội dung tài liệu. Có thể bao gồm tóm tắt, chú thích, mục lục, đoạn văn bản để làm rõ nội dung...
  5. Nhà xuất bản (Publisher): Nhà xuất bản, nơi ban hành tài liệu có thể là tên cá nhân, tên cơ quan, tổ chức, dịch vụ...
  6. Tác giả phụ (Contributor): Tên những người cùng tham gia cộng tác đóng góp vào nội dung tài liệu, có thể là cá nhân, tổ chức...
  7. Ngày tháng (Date): Ngày, tháng ban hành tài liệu.

8. Loại (kiểu) (Type): Mô tả bản chất của tài liệu. Dùng các thuật ngữ mô tả phạm trù kiểu: trang chủ, bài báo, báo cáo, từ điển...
9. Khổ mẫu (Format): Mô tả sự trình bày vật lý của tài liệu, có thể bao gồm; vật mang tin, kích cỡ độ dài, kiểu dữ liệu (.doc,.html,.jpg, xls, phần mềm....)
10. Định danh (Identifier): Các thông tin về định danh tài liệu, các nguồn tham chiếu đến, hoặc chuỗi ký tự để định vị tài nguyên: URL (Uniform Resource Locators) (bắt đầu bằng http://), URN (Uniform Resource Name), ISBN (International Standard Book Number), ISSN (International Standard Serial Number), SICI (Serial Item & Contribution Identifier),...
11. Nguồn (Resource): Các thông tin về xuất xứ của tài liệu, tham chiếu đến nguồn mà tài liệu hiện mô tả được trích ra/tạo ra, nguồn cũng có thể là: đường dẫn (URL), URN, ISBN, ISSN...
12. Ngôn ngữ (Language): Các thông tin về ngôn ngữ, mô tả ngôn ngữ chính của tài liệu.
13. Liên kết (Relation): Mô tả các thông tin liên quan đến tài liệu khác. có thể dùng đường dẫn (URL), URN, ISBN, ISSN...
14. Diện bao quát (Coverage): Các thông tin liên quan đến phạm vi, quy mô hoặc mức độ bao quát của tài liệu. Phạm vi đó có thể là địa điểm, không gian hoặc thời gian, tọa độ...
15. Bản quyền (Right): Các thông tin liên quan đến bản quyền của tài liệu.



- **DESIGN:** Dùng để thiết kế giao diện bộ sưu tập kèm theo những chỉ thị tìm kiếm theo những dẫn mục cho ta chọn, chẳng hạn như: tác giả, nhan đề, năm,...
- **CREAT:** Dùng để sản xuất bộ sưu tập.
- **FORMAT:** Trình bày tài liệu.

#### *b. Các bước xây dựng bộ sưu tập với GLI*

- Bước 1: Khởi động GLI  
Start/ Program/ Greenstone/Greenstone Librarian Interface.  
Chọn File/ New để tạo bộ sưu tập mới: Gõ tên vào mục Collection title và mô tả tóm tắt vào phần Description of content.
- Bước 2: Chọn tài liệu cần tạo bộ sưu tập  
Bấm Gather/ Mở đến địa chỉ chứa tài liệu trong phần workspace.  
Chọn tập tin tài liệu kéo và thả vào cửa sổ Collection bên phải.
- Bước 3: Biên mục cho từng tài liệu đưa vào theo chuẩn Dublin Core  
Bấm vào Enrich/ Chọn file tương ứng rồi biên mục theo 15 trường của Dublin Core.
- Bước 4: Xây dựng bộ sưu tập  
Create/ Build collection.

#### **3.2.2. Xây dựng bộ sưu tập Luận văn tốt nghiệp**

### **3.3. THIẾT KẾ GIAO DIỆN MỘT SỐ TRANG LIBWEB**

#### **3.4. THỬ NGHIỆM**

##### **3.4.1. Công cụ và môi trường phát triển hệ thống**

##### **3.4.2. Đánh giá**

## KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

### 1. Kết luận

GreenStone là hệ thống thư viện số đơn giản, nhưng không kém phần hiệu quả trong việc quản lý tài nguyên điện tử. Các thư viện số xây dựng bởi GreenStone tuân thủ theo các tiêu chuẩn quốc tế chung, giúp cho tiến trình liên kết các thư viện, vốn đang là xu hướng hiện nay, trở nên dễ dàng hơn. Đặc biệt, GreenStone hoàn toàn miễn phí và được người dùng trên toàn thế giới sử dụng và đóng góp cho sự phát triển. Những yếu tố này rất có ý nghĩa với những thư viện hạn chế về năng lực tài chính như ở Việt Nam.

#### ❖ Kết quả đạt được

Chúng tôi đã thực hiện được những mục tiêu đặt ra ban đầu: tìm hiểu cấu trúc và cơ chế hoạt động của GreenStone, từ đó, xây dựng ứng dụng thư viện số cho Trường Đại học Phú Yên. Hệ thống ngoài việc thực hiện được các chức năng cơ bản như GreenStone (tìm kiếm thông tin tài liệu, tra cứu trên nội dung tài liệu) còn bổ sung thêm các tính năng riêng, xây dựng các bộ sưu tập đặc thù, giao diện tương tác với người dùng đẹp, dễ sử dụng.

Tuy nhiên, hệ thống còn hạn chế trong việc chưa cung cấp đầy đủ các chức năng: các dịch vụ trực tuyến, diễn đàn, đăng ký thành viên, thể loại và số lượng các bộ sưu tập còn ít.

### 2. Hướng phát triển

GreenStone là phần mềm mã nguồn mở, khả năng phát triển trong tương lai là rất lớn.

Trong hệ thống LibWeb cần bổ sung các chức năng khác: các dịch vụ trực tuyến, diễn đàn, đăng ký thành viên, phân phát tự động tài liệu đến người đăng ký sử dụng (theo ý thích của họ).