ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC XÃ HỘI VÀ NHÂN VĂN
THƯ VIỆN TRƯỜNG

QUY TRÌNH
SỐ HÓA TÀI LIỆU

MÃ SỐ : **TV.QT20.01**
LẦN BAN HÀNH : **01**
NGÀY BAN HÀNH : **24/08/2009**

SOẠN THẢO	KIỂM TRA	PHÊ DUYỆT
Nguyễn Thị Xuân Anh		PGS.TS.Võ Văn Sen
Thư viện trường	Ban Cải tiến	Hiệu trưởng

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC XÃ HỘI VÀ NHÂN VĂN

NHỮNG THAY ĐỔI ĐÃ CÓ

Ngày tháng	Nội dung thay đổi	Lần ban hành
24/08/2009	Ban hành lần đầu	01

1. MỤC ĐÍCH:

Quy định các bước thực hiện số hóa tài liệu của thư viện.

2. PHẠM VI ÁP DỤNG

Quy trình áp dụng cho tất cả cán bộ thư viện làm công tác số hóa tài liệu.

3. ĐỊNH NGHĨA:

- TV: Thư viện
- CB: Cán bộ
- CBTV: Cán bộ thư viện
- NSD: Người sử dụng

4. TÀI LIỆU THAM KHẢO:

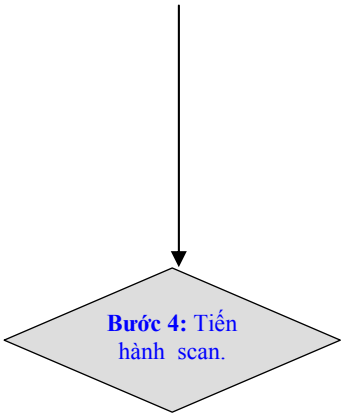
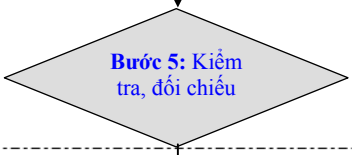
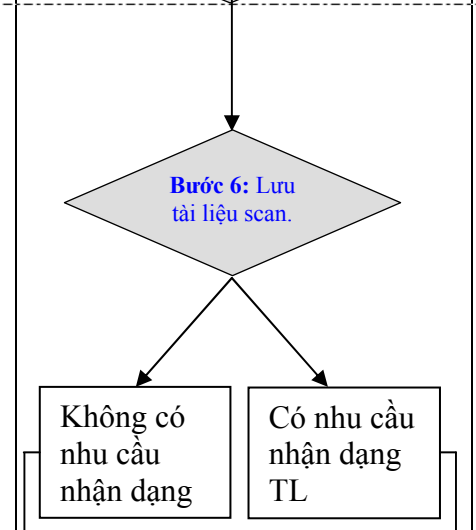
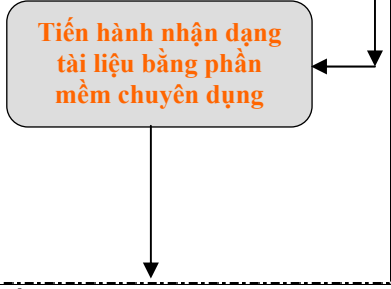
Quy định về chức năng, nhiệm vụ của TV Trường.

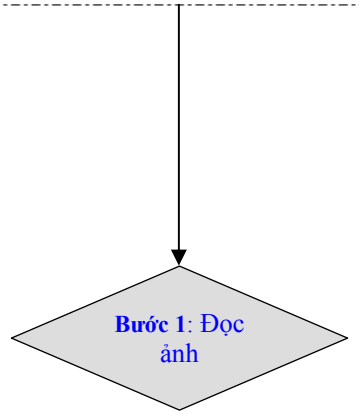
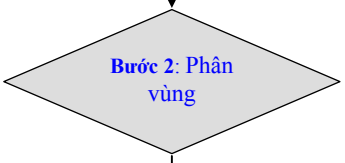
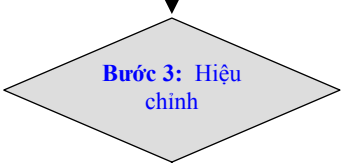
5. NỘI DUNG:

5.1. Quy trình nghiệp vụ:

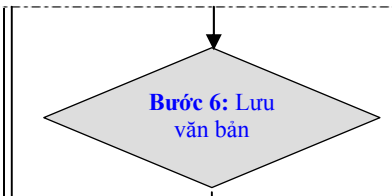
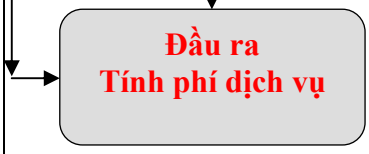
Người thực hiện	Quy trình thực hiện	Mô tả công việc
CBTV	**Đầu vào**	❖ Tiêu chuẩn đầu vào: – Chọn tài liệu cần scan – Tài liệu giấy đã được xử lý kỹ thuật trước khi scan. – Chuẩn bị trang thiết bị: Xem xét nguồn điện, ổn áp, khởi động máy tính, máy scan, ánh sáng, … – Thiết lập, quy định các thông số trên phần mềm dựa trên loại tài liệu cần số hóa. – Xác định dạng lưu trữ theo yêu cầu của tài liệu được số hóa.

	Lựa chọn tài liệu số hóa hoặc nhận yêu cầu số hóa	❖ 5 tiêu chí áp dụng để lựa chọn tài liệu số hóa: – Tình trạng bản quyền của tài liệu. – Nhóm người sử dụng mà thư viện xác định mức độ ưu tiên phục vụ. – Nội dung tài liệu. – Điều kiện bảo quản hiện tại – Các loại tài liệu đặc biệt. – …
CB số hóa	**Tiến hành scan tài liệu bằng máy chuyên dụng**	
CB số hóa	**Bước 1**: Khởi động máy scan	❖ Bật nút công tắc ***on/off*** của máy scan
CB số hóa	**Bước 2**: Mở phần mềm scan	❖ Click chuột vào biểu tượng ***HP Solution Center.***
CB số hóa	**Bước 3:** Settings	❖ Thiết lập các chế độ chuẩn trước khi tiến hành scan. – Chọn nút ***Setting*** ➡ ***Scan Setting and Preferences*** ➡ ***Button Settings.*** – Chọn chế độ scan: + Chọn ***Scan Picture:*** đối với scan tài liệu để đưa vào CSDL toàn văn và scan dịch vụ không xuất ra file text; + Chọn ***Scan Document:*** đối với scan dịch vụ xuất ra file text. – Chọn độ phân giải : 300ppi, Black and White. – Chọn nơi lưu trữ : Tại nút ***Save Option*** ➡ *Chọn nơi muốn lưu* (Thông thường máy tự động để đường dẫn vào ***My Document / My Scan***/…), sau đó ***click OK***.

CB số hóa		❖ Có 2 cách scan **:** – Cách 1: Nhấn phím ***Delete*** trên bàn phím để bỏ khung ban đầu và kẻ khung bằng tay (thủ công). – Cách 2: + Tạo 1 khung mẫu rồi vào ***Scan/ Save***. Đặt tên tùy ý rồi click ***Save*.** + Lần sau mỗi khi scan chỉ việc vào ➡ **Scan ➡ Load ➡ Ok.** – Mục đích của việc tạo khung mẫu để các trang trong cùng 1 tài liệu có cùng kích cỡ và đều nhau. – Khi chỉnh khung xong click vào nút ***Accept.*** – Chọn **Yes** để tiếp tục scan trang mới, chọn **No** để **Stop** scan và **Save** lại.
CB số hóa		❖ Kiểm tra, đối chiếu với bản gốc để chỉnh sửa lỗi tài liệu vừa scan.
CB số hóa		❖ **Lưu ý :** – Scan từng trang đối với tài liệu lớn, và có thể scan 2 trang đối với tài liệu nhỏ. – Scan từng trang phải chỉnh (quay) tài liệu cho đúng chiều của nó, tránh trường hợp scan ngược. – Khi scan xong, sau khi kiểm tra chưa đạt yêu cầu thì **delete** trang đó và scan lại, máy sẽ tự động lưu vào vị trí file đã xóa. – Nên scan xong tài liệu rồi mới tiến hành chỉnh sửa, không nên vừa scan vừa chỉnh sửa sẽ tốn nhiều thời gian và không chính xác. – Khi đã scan khoảng 20 - 30 trang cần lưu lại bằng cách chọn **No** và thực hiện lại bước 4.
CB số hóa		❖ Tiêu chuẩn file nhận dạng: – File ảnh dạng *.TIF – Ảnh đen / trắng – Độ phân giải 300 DPI trở lên, 400 DPI đối với văn bản có cỡ chữ dưới 8

CB số hóa		❖ Khởi động phần mềm: – Bấm vào biểu tượng **VnDOCR** trên **Desktop** – Hoặc vào **Start** → **Programs** → **VnDOCR** ❖ Mở các file ảnh cần nhận dạng. ❖ Thao tác: – Bấm vào biểu tượng **Đọc ảnh** trên thanh công cụ. – Hoặc vào Menu **Tệp** → chọn **Đọc ảnh**. – Chọn các file ảnh cần nhận dạng → Bấm **Mở** **Kết quả:** Các file ảnh cần nhận dạng đã được đưa vào phần mềm.
CB số hóa		❖ Lựa chọn và thiết lập các vùng nhận dạng trên từng trang văn bản. ❖ Thao tác: – **Phân vùng tự động**: đối với các trang văn bản có cấu trúc đơn giản + Bấm biểu tượng **Phân vùng** trên "Thanh công cụ chính" + Hoặc chọn **Tự động phân vùng** trong Menu **Xử lý**. – **Phân vùng bằng tay**: Đối với trang văn bản có cấu trúc phức tạp, trang cần phải sửa lại một số vùng, thực hiện như sau: + Bấm vào biểu tượng **Tạo vùng mới** trên thanh công cụ; + Hoặc vào Menu **Xử lý** → chọn **Đánh dấu vùng bằng tay** → chọn **Tạo mới** Sau đó thiết lập cho các vùng khác nhau trên từng trang văn bản, phần mềm hỗ trợ 3 định dạng: văn bản, bảng biểu, ảnh. **Kết quả:** thiết lập chính xác các vùng nhận dạng (văn bản, bảng biểu, ảnh)
CB số hóa		❖ Bước này được thực hiện trong trường hợp file ảnh cần nhận dạng có chất lượng không tốt như nét chữ mờ hoặc đậm quá, ảnh bị nghiêng… ❖ Thao tác (thực hiện riêng lẻ cho từng trang văn bản), bao gồm: – Chỉnh ảnh bị nghiêng: vào Menu **Công cụ** → chọn **Xác định độ nghiêng** – Các tính năng khác: Vào Menu **Công cụ**

		→ chọn **Công cụ ảnh** → Chọn tính năng cần thao tác (**Làm dày nét, Làm mảnh nét, Tăng chất lượng ảnh**…) Sau khi hiệu chỉnh trang nào cần phải phân vùng lại trang đó. **Kết quả:** tăng chất lượng các file, đảm bảo nhận dạng chính xác hơn.
CB số hóa	**Bước 4:** Nhận dạng	❖ Chuyển các file ảnh sau khi đã được phân vùng (và hiệu chỉnh nếu có) thành tài liệu dạng text. ❖ Thao tác: – Bấm vào biểu tượng **Nhận dạng** trên thanh công cụ chính; – Hoặc vào Menu **Xử lý** → chọn **Nhận dạng** → xuất hiện hộp thoại **Miền tác động,** chọn 1 trong 3 lựa chọn sau: + **Tất cả** (Nhận dạng tất cả các trang); + **Trang hiện thời** (chỉ nhận dạng trang đang hiển thị); + **Trang** (Chính xác số trang cần nhận dạng) → Rồi bấm **Đồng ý** **Kết quả:** tài liệu được chuyển sang dạng text (chưa hoàn chỉnh).
CB số hóa	**Bước 5:** Kiểm tra	❖ Sau khi nhận dạng cần kiểm tra lại kết quả so với văn bản gốc để sửa những lỗi chính tả. ❖ Thao tác: – Bấm vào biểu tượng **Chính tả** trên thanh công cụ chính; – Hoặc nhấn phím tắt **F7**; – Hoặc vào Menu **Công cụ** → chọn **Soát chính tả** → Xuất hiện hộp thoại **Soát chính tả** hướng dẫn chi tiết các từ sai, lỗi cần sửa. Lưu ý màu sắc của các từ báo những tín hiệu khác nhau (Ví dụ: Từ nghi ngờ**:** màu xanh ; Từ sai**:** màu đỏ) → Sửa chính tả các từ sai sau khi đã đối chiếu với vản bản gốc cửa sổ phía dưới. Trong quá trình kiểm tra, nếu chất lượng file nhận dạng không tốt có thể quay lại các bước (2), (3), (4). **Kết quả:** Tài liệu dạng text hoàn chỉnh.
CB số hóa		❖ Ghi kết quả nhận dạng ra tệp. ❖ Thao tác: – Phím tắt **Ctrl+S**; – Hoặc vào Menu **Tệp** → chọn **Lưu văn**

<table>
<tr>
<td></td>
<td> 
</td>
<td>bản → Hộp thoại xuất hiện → Xác lập tên, nơi lưu trữ, định dạng thích hợp → bấm Lưu
<u>Kết quả</u>: Hoàn thành chuyển dạng tài liệu từ File ảnh sang dạng text.</td>
</tr>
<tr>
<td>CB số hóa</td>
<td> 
</td>
<td>❖ Tiêu chuẩn đầu ra đối với file scan:
– File ảnh dạng *.TIF, .pdf, .jpeg, …
– Ảnh đen/trắng (đối với nội dung tài liệu), màu (đối với hình ảnh, trang bìa)
– Độ phân giải 300 PPI trở lên, 400 PPI đối với văn bản có cỡ chữ dưới 8.
– Các file ảnh của tài liệu phải đầy đủ, rõ ràng và dễ đọc.
❖ Tiêu chuẩn đầu ra đối với file nhận dạng: Tùy theo yêu cầu đầu ra, phần mềm hỗ trợ các định dạng:
– Microsoft Word 7.0 (*.doc)
– Microsoft Word 2000-XP (*.doc)
– Microsoft Word 7.0 – RichTextFormat (*.rtf)
– Microsoft Word 2000-XP – RichTextFormat (*.rtf)
– Text Only
Tính phí dịch vụ bao gồm scan và nhận dạng.</td>
</tr>
</table>