

SỐ HÓA VỚI HỆ THỐNG KIRTAS

Đỗ Như Thơ và Trần Đức Trung
Cục Thông tin KH&CN Quốc gia

Giới thiệu kinh nghiệm của Cục Thông tin KH&CN Quốc gia (NASATI) trong khai thác các thiết bị Kirtas để số hóa tài liệu.

Ngày nay, công nghệ thông tin và kỹ thuật máy tính phát triển không ngừng đã thúc đẩy quá trình biến đổi các tài liệu in trên giấy thành tài liệu dạng điện tử thuận tiện trong sử dụng. Quá trình biến đổi này, nhờ sự phát triển mạnh mẽ của công nghệ, trở nên đơn giản hơn rất nhiều. Bài viết này đề cập đến các kinh nghiệm trong vận hành, khai thác một trong số các thiết bị số hóa hiện đại - thiết bị của

hãng Kirtas (Hoa Kỳ).

1. Vài nét về thiết bị số hóa Kirtas

Nhu cầu tăng tốc độ số hóa các tài liệu in truyền thống phần nào được đáp ứng nhờ thiết bị số hóa của hãng Kirtas (Mỹ). Thiết bị Kirtas APT1600 và Kabis (Hình 1) được thiết kế phù hợp với những đòi hỏi về số hóa tài liệu đóng tập. Về cơ bản, hệ thống vận hành và công suất thiết bị của Kirtas APT1600 và Kabis không khác nhau nhiều.



Hình 1. Thiết bị Kirtas AP1600 và Kabis

Toàn bộ hệ thống số hóa của hãng Kirtas đều sử dụng máy ảnh kỹ thuật số chất lượng cao, khác với công nghệ máy quét phẳng truyền thống để sao chụp hình ảnh. Một máy ảnh kỹ thuật số Canon EOS-1Ds Mark II, 16,7 triệu điểm ảnh,

cho phép chụp liên tục mỗi trang trái và trang phải của cuốn sách nhờ sử dụng một gương phản chiếu qua ống kính camera EF tiêu cự 24-70 mm. Độ phân giải có thể cài đặt ở mức 300 ppi (pixel per inch) và có khả năng tăng đến 600 ppi. Đồng thời,

Nghiên cứu - Trao đổi

công nghệ lật trang tự động SureTurn™ cũng giúp giảm bớt thao tác lật trang bằng tay, (thường không được hoan nghênh đối với những tài liệu đóng tập cũ và dễ hỏng rách). Với cánh tay rôbot được điều khiển bằng máy tính sử dụng một hệ thống hút chân không, hệ thống page luffers nhẹ nhàng lật chỉ một trang duy nhất tại một thời điểm. Cùng thời điểm đó, cuốn sách nhờ bộ phận nâng sách được thiết kế theo công nghệ SmartCradle được nhẹ nhàng nâng lên, tự động dịch chuyển để bù vào việc giảm số lượng trang sách bên phải khi chúng được lật giở và sao chụp. Công nghệ SmartCradle giữ sách mở một góc đúng bằng 110 độ trong suốt quá trình vận hành, tạo ra một môi trường ứng suất thấp để xử lý những sách quý hiếm và dễ hỏng rách.

Để đảm bảo việc chỉ lật duy nhất một trang, thiết bị số hóa được thiết kế với bộ cảm biến Page Edge Sensor và thiết bị chia tách trang Page Separator. Thiết bị Page Separator giúp chia tách trang khi bộ cảm biến Page Edge Sensor giám sát việc vận hành máy với độ chính xác cao và điều chỉnh đầu trang sách khi cần. Đồng thời, nó liên tục kiểm tra nhằm đảm bảo cánh tay rôbot chỉ lật một trang duy nhất tại một thời điểm, và kiểm soát từ xa khi có trang bị bỏ sót, khi đó, bộ cảm biến Page Edge Sensor sẽ tạm ngừng vận hành máy để con người thực hiện thao tác điều chỉnh. Kirtas APT 1600 và Kabis có thể sao chụp đến 1600 trang một giờ, nhưng đó chưa phải là đặc điểm duy nhất về năng suất của máy.

Còn phải kể đến phần mềm vận hành hệ thống số hóa dễ dàng thao tác và thuận tiện cho người sử dụng; cùng với sự hỗ trợ của chức năng lật trang, mỗi bước thiết lập và vận hành máy đều được chỉ dẫn thông qua những đoạn video hoàn chỉnh. Các tiện ích này luôn sẵn có cho tất cả khách hàng sử dụng hệ thống Kirtas thông qua Mạng nội bộ khách hàng Kirtas (Kirtas Customer Extranet), nơi cho phép dễ dàng tải những phiên bản cập nhật phần mềm và tài liệu mới.

2. Thông số kỹ thuật của thiết bị Kirtas

- Tốc độ quét: 1600 trang/giờ.
- Hình ảnh: màu, đen/trắng.
- Định dạng đầu ra: TIFF, JPEG, RAW,...
- Sao chụp: máy ảnh kỹ thuật số 16.7 triệu điểm ảnh, độ sâu 24 bit RGB, 4,992 x 3,228 ảnh điểm.
- Xử lý ảnh: độ phân giải 300 ppi có thể tăng lên đến 600 ppi.
- Kích cỡ trang giấy: 11,4 cm x 17,8 cm đến 27.9 cm x 35,5 cm.
- Độ dày trang giấy: 20 g/m² đến 300 g/m²
- Độ dày đóng tập: 10,16 cm.
- Kích cỡ máy: 84 cm x 76 cm x 122 cm (Rộng x Dài x Cao).
- Tính chính xác: Bộ cảm biến Page Edge Sensor.
- Cải tiến: Thiết bị chia tách trang Page Separator.
- Trọng lượng: 73 Kg.

Nghiên cứu - Trao đổi

- Nguồn điện: 110-240 VAC, 50/60 Hz, 15 A.

- Trạm xử lý: Máy chủ Kirtas Image Server.

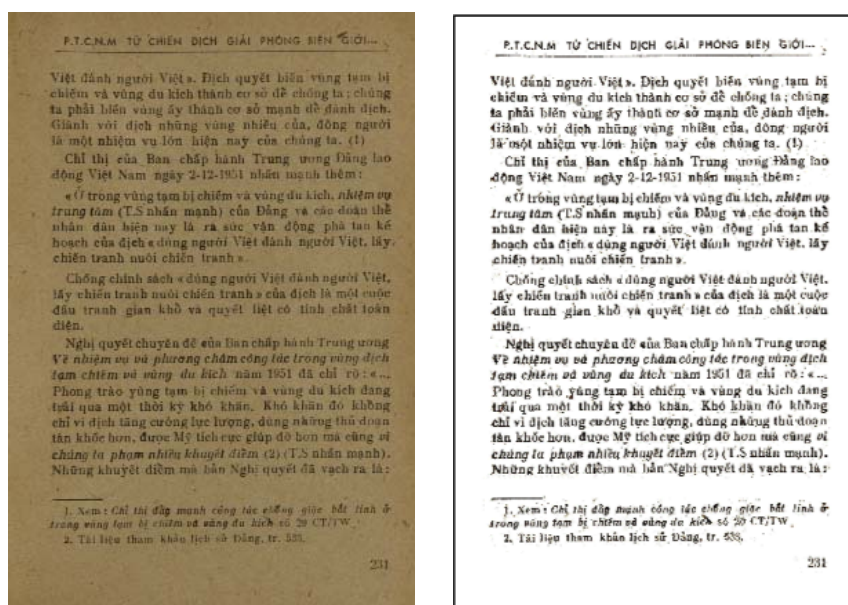
Ngoài ra, hệ thống số hóa Kirtas còn tích hợp phần mềm BookScan Editor cho phép căn chỉnh trang tự động theo khối dữ liệu với năng suất cao và phần mềm nhận dạng ký tự quang học (OCR) cho 189 ngôn ngữ.

3. Kinh nghiệm vận hành, khai thác

Gần hai năm sử dụng thiết bị Kirtas để số hóa tài liệu tại Cục Thông tin KH&CN Quốc gia (NASATI) cho phép

chúng tôi rút ra một số nhận xét sau đây:

- Có thể nói, hệ thống số hóa Kirtas thực sự là lý tưởng cho những ấn phẩm đạt tiêu chuẩn thông thường. Đối với một số tài liệu in trên giấy quá mỏng như giấy Poluya, thì việc số hóa sẽ được tiến hành thông qua chức năng chụp bằng tay cho từng trang và việc căn chỉnh sau đó sẽ được thực hiện như thông thường bằng phần mềm BookScan Editor (như đã nói ở trên). Hệ thống số hóa này cho phép phục hồi phần nào những tài liệu sử dụng giấy quá mỏng hay quá ngả màu theo thời gian (Hình 2).



Hình 2. Tài liệu trước và sau khi số hóa với hệ thống Kirtas.

- Kirtas được thiết kế với độ rộng của gáy sách hay độ dày đóng tập ở mức tiêu chuẩn để phù hợp và tương ứng với hệ thống hơi thổi từ phía thành máy, để trong quá trình số hóa, từng trang sách được thổi toi giúp cho cánh tay lật trang

hoạt động dễ dàng. Do đó, nếu sách cần số hóa quá mỏng thì nó phải được đặt trên một quyển sách khác ở bên phải trang sách để tăng độ dày đóng tập đến mức cho phép hệ thống hơi thổi và cánh tay lật trang hoạt động.

Nghiên cứu - Trao đổi

- Trong quá trình số hóa, người vận hành cần quan sát và theo dõi sự liên tục theo trật tự của số trang (tránh bị đúp trang). Khi thấy trang sách quá phồng, cần dùng tay vuốt nhẹ để hai trang sách trải đều sang hai phía, như vậy việc số hóa sẽ đạt chất lượng tốt nhất.

- Đối với sách có độ dày gáy lớn, sau khoảng 50 - 60 trang sách, cần dừng máy ở chế độ tạm thời để máy tự động căn chỉnh tâm của gáy sách vào chính giữa hai khe đỡ giá sách.

- Cần quan tâm đến độ rộng của tài liệu số hóa để điều chỉnh máy cho hợp lý

và chính xác. Việc này nên được làm ngay từ đầu cho các tài liệu có cùng độ rộng trong cùng một thời gian máy hoạt động để tiết kiệm thời gian căn chỉnh máy.

Số hóa tài liệu, đặc biệt là các di sản văn hóa, đang là vấn đề thời sự của hoạt động thông tin-thư viện được nhiều nước trên thế giới quan tâm. Việc triển khai hiệu quả công việc này phụ thuộc nhiều vào sự lựa chọn thiết bị số hóa, và kèm theo đó là, việc đào tạo con người. Hy vọng kinh nghiệm của chúng tôi sẽ góp phần giải quyết bài toán nói trên.

CÁC SẢN PHẨM NĂM 2011 CỦA CỤC THÔNG TIN KH&CN QUỐC GIA

1. **Tạp chí Thông tin và Tư liệu:** là tạp chí hàng đầu của ngành Thông tin-Tư liệu-Thư viện-Thống kê KH&CN. Giới thiệu các kết quả nghiên cứu khoa học, kinh nghiệm triển khai hoạt động thông tin KH&CN, các vấn đề khoa học chuyên ngành đang được thế giới quan tâm. Cung cấp thông tin về các sự kiện, văn bản pháp quy liên quan đến hoạt động của toàn bộ mạng lưới Thông tin-Tư liệu-Thư viện-Thống kê KH&CN cả nước.

48 trang - 4 số/năm - 15.000đ/số

2. **Khoa học-Công nghệ-Môi trường:** cung cấp thông tin về dự báo, chiến lược, chính sách phát triển KH&CN trong và ngoài nước và về đổi mới công nghệ. Giới thiệu các thành tựu KH&CN nổi bật, thông tin chọn lọc về môi trường, các sự kiện lớn trong lĩnh vực KH&CN và môi trường.

48 trang - 12 số/năm - 15.000đ/số

3. **Khoa học và Công nghệ Thế giới các năm 2002, 2004, 2005, 2006 và 2007:** phân tích xu thế phát triển, tổng kết các chính sách và năng lực KH&CN của các nước trên thế giới.

300-400 trang - 50.000đ/quyển

4. **Kết quả thực hiện nhiệm vụ khoa học và công nghệ:** cung cấp thông tin về kết quả thực hiện các đề tài, dự án: thuộc Chương trình KH&CN trọng điểm cấp Nhà nước, độc lập cấp Nhà nước và cấp Bộ được thực hiện trong năm. Thông tin này bao gồm: tên đề tài, chủ nhiệm, cơ quan chủ trì, cơ quan chủ quản, cấp, thời gian hoàn thành, tóm tắt kết quả, nơi lưu trữ báo cáo kết quả

80-120 trang - 2 số/năm - 25.000đ/số

5. **Tổng luận Khoa học-Công nghệ-Kinh tế:** cung cấp thông tin về các vấn đề nổi bật trong lĩnh vực kinh tế-xã hội, KH&CN, môi trường dưới dạng tổng luận hoặc chuyên khảo.

36-60 trang - 12 số/năm - 20.000đ/số

6. **Nhiệm vụ KH&CN đang tiến hành:** cung cấp thông tin về các đề tài, dự án đang được thực hiện trong năm: thuộc chương trình KH&CN trọng điểm cấp Nhà nước, độc lập cấp Nhà nước và cấp Bộ. Thông tin này bao gồm: tên đề tài, chủ nhiệm, cơ quan chủ trì, cơ quan chủ quản, cấp, thời gian bắt đầu, thời gian kết thúc dự kiến, tóm tắt nội dung, nơi lưu trữ bản thuyết minh.

100-150 trang - 2 số/năm - 25.000đ/số