CHƯƠNG 4
XỬ LÝ SỐ LIỆU VÀ PHỔ BIẾN KẾT QUẢ

Để đẩy nhanh tốc độ xử lý và nâng cao chất lượng số liệu Tổng điều tra, một chiến lược quan trọng cần được thử nghiệm và kết luận sớm là đổi mới công nghệ xử lý theo công nghệ nhận biết ký tự thông minh (công nghệ quét). Đây là công nghệ tiên tiến, lần đầu tiên đưa vào áp dụng trong điều tra thống kê ở nước ta nên công tác chuẩn bị phải tiến hành hết sức thận trọng và tỉ mỉ.

Nhận thức được vấn đề nói trên, ngay từ năm 2006, thông qua Dự án VNM7PG0008, Trung tâm Tin học Thống kê Khu vực I đã được trang bị một máy quét Fujitsu và phần mềm quét Readsoft (Thụy Điển) để tiến hành thử nghiệm. Phiếu của các cuộc Điều tra biến động dân số và kế hoạch hoá gia đình các năm 2007 và 2008, và phiếu của một điều tra thí điểm và điều tra tổng duyệt của Tổng điều tra đã được sử dụng để thử nghiệm và làm quen với công nghệ quét từ mức độ nhỏ (ít phiếu điều tra) đến lớn (nhiều phiếu điều tra). Các thử nghiệm này cho phép đánh giá năng lực cán bộ tin học trong việc ứng dụng công nghệ nhận biết ký tự thông minh, các yêu cầu kỹ thuật liên quan như bút viết, loại giấy sử dụng in phiếu, máy cắt, kho chứa phiếu, v.v…

Trong quá trình thử nghiệm, Lãnh đạo Tổng cục Thống kê đã chỉ đạo sát sao Vụ Thống kê Dân số và Lao động, Văn phòng Ban Chỉ đạo Trung ương, Trung tâm Tin học Thống kê Khu vực I và Dự án "Hiện đại hóa Tổng cục Thống kê" chuẩn bị kỹ lưỡng mọi điều kiện cần thiết để đảm bảo cho việc áp dụng thành công, như: thiết kế phiếu, quy cách giấy in phiếu, in phiếu, bút viết và các văn phòng phẩm khác, các phương tiện bảo quản phiếu, cơ sở hạ tầng về công nghệ thông tin, phần mềm và các máy móc thiết bị khác cho các Trung tâm xử lý số liệu. Trong thử nghiệm này, Tổng cục Thống kê đã nhận được hỗ trợ kỹ thuật quý báu từ Cục Thống kê Úc thông qua việc cử chuyên gia giàu kinh nghiệm về xử lý thông tin sang Việt Nam công tác.

4.1 Giới thiệu

Thông qua gói thầu GSO6.1B với Liên danh nhà thầu FPT[4]-TIS[5], Dự án "Hiện đại hóa Tổng cục Thống kê" đã trang bị hệ thống xử lý hiện đại cho 3 Trung tâm Tin học Thống kê với tổng số 25 máy chủ, 149 máy tính cá nhân, các phần mềm hệ thống và phần mềm xử lý TIS eFlow cùng với nhiều thiết bị và phần mềm khác, nâng cấp mạng cục bộ của các Trung tâm, đảm bảo những điều kiện tốt nhất cho công tác xử lý (xem Phụ lục 19).

Mặc dù việc triển khai cung cấp thiết bị, phần mềm và xây dựng hệ thống xử lý bị chậm trễ hơn dự kiến, đến đầu tháng 8/2009 các Trung tâm Tin học Thống kê mới bắt đầu triển khai xử lý phiếu Tổng điều tra. Nhưng với kinh nghiệm có được trong việc thử nghiệm trình bày ở trên và tinh thần phấn đấu quyết liệt, đến hết tháng 10/2009 việc nhập số liệu của phiếu mẫu (bao gồm việc quét phiếu hoặc nhập tin, nhận dạng và kiểm tra nhận dạng, kiểm tra logic sơ bộ) đã hoàn thành. Tỷ lệ phiếu được nhận dạng đạt trên 95%. Đây là kết quả ấn tượng so với dự kiến ban đầu.

Trong tháng 11-12/2009, Trung tâm Tin học Thống kê Khu vực I phối hợp với Vụ Thống kê Dân số và Lao động tiến hành kiểm tra và hiệu chỉnh số liệu mẫu 15% và tổng hợp các biểu kết quả. Một số chỉ tiêu tổng hợp chính đã được công bố vào ngày 31/12/2009 theo

4 Công ty Cổ phần FPT của Việt Nam.

5 Công ty TIS (Top Image System). Đây là công ty của Israel chuyên về giải pháp nhập tin một cách chính xác, hiệu quả và tự động.

đúng kế hoạch. Việc kiểm tra, đánh giá các kết quả tổng hợp số liệu suy rộng mẫu 15% đã được tiến hành khẩn trương và giữa tháng 2/2010 đã phân phát dữ liệu vi mô cho các Cục Thống kê và các cơ quan sử dụng thông tin khác.

Công tác nhập liệu phiếu điều tra toàn bộ đã được tiến hành khẩn trương, thận trọng và hoàn thành vào tháng 4/2010. Tuy nhiên, so với thời gian dự kiến hoàn thành khâu quét và hoàn thiện kết quả nhập tin phiếu điều tra toàn bộ thì tiến độ đã bị chậm 2 tháng do hầu hết máy quét (scanners) đã bị các sự cố như kẹt giấy, dãn ảnh, co ảnh,… Nguyên nhân chính là do bánh xe cuốn và thanh thép bị mòn nhưng không có thiết bị thay thế kịp thời, việc xử lý sự cố của nhà thầu khá chậm, ảnh hưởng đến tiến độ xử lý.

Tiến độ hoàn thành công tác xử lý phiếu điều tra mẫu 15% và điều tra toàn bộ cho phép khẳng định rằng việc áp dụng công nghệ nhận biết ký tự thông minh đã rút ngắn thời gian và nhân lực nhập tin. Đây là kết quả rất đáng khích lệ, củng cố niềm tin vào việc lựa chọn giải pháp công nghệ này cho các cuộc Tổng điều tra và điều tra mẫu khác trong thời gian tới.

Ngoài việc áp dụng công nghệ quét để xử lý phiếu, trong Tổng điều tra 2009, công nghệ thông tin cũng được áp dụng nhiều hơn và sâu hơn như: tin học hóa quy trình quản lý phiếu tại các Cục Thống kê trong giao nhận, ghi mã, kiểm tra, lập bảng kê, chọn mẫu, tổng hợp sơ bộ, v.v… Tại các Trung tâm Tin học Thống kê, toàn bộ quy trình xử lý đều được tin học hóa: quá trình giao nhận, chuẩn bị phiếu, truyền tin và dữ liệu, theo dõi luồng phiếu và luồng dữ liệu, kiểm soát số lượng,… Trang Web tác nghiệp Tổng điều tra cũng đã phát huy tác dụng trong việc trao đổi thông tin, hỏi đáp nghiệp vụ, báo cáo tiến độ giữa Ban Chỉ đạo Trung ương, Ban Chỉ đạo tỉnh, thành phố và các Trung tâm Tin học Thống kê tham gia xử lý số liệu Tổng điều tra.

Đây là lần đầu tiên Tổng cục Thống kê áp dụng công nghệ nhận biết ký tự thông minh dạng trên diện rộng. Hơn nữa hệ thống được thiết kế và xây dựng trong một thời gian ngắn nên quá trình xử lý còn nhiều vướng mắc và hạn chế. Cụ thể là như sau:

- Gói thầu của Dự án "Hiện đại hoá Tổng cục Thống kê" được triển khai chậm hơn dự kiến nên việc cung cấp thiết bị, phần mềm bị chậm, hệ thống xử lý được xây dựng và phát triển trong thời gian quá ngắn không đủ thời gian thử nghiệm. Hơn nữa, phần mềm TIS eFlow hoàn toàn mới đối với Tổng cục Thống kê (trước đó sử dụng phần mềm ReadSoft cho thử nghiệm trong hai năm 2007-2008). Vì thế, hệ thống xử lý còn thiếu các chương trình ứng dụng hoàn hảo ngay từ đầu mà phải qua nhiều lần điều chỉnh, sửa chữa, hoàn thiện, làm cho quá trình xử lý bị động. Các cuộc điều tra sau này cần dành nhiều thời gian hơn cho việc xây dựng, phát triển ứng dụng, nhất là cho việc kiểm tra và chạy thử.

- Giấy in và chất lượng in đặc biệt ảnh hưởng đến thời gian, chi phí lao động và độ chính xác của việc xử lý. Giấy in phiếu cũng như việc in phiếu cũng cần được rút kinh nghiệm và nâng cao chất lượng hơn nữa. Đặc biệt là với phần mềm xử lý phiếu TIS đặt ra những yêu cầu khác biệt, cao hơn so với phần mềm ReadSoft đã sử dụng trong các thử nghiệm trước đây.

- Việc ghi phiếu cũng cần được đánh giá, rút kinh nghiệm để khắc phục tối đa những lỗi ghi sai như: viết đảo số, ghi sai cột, viết chữ số quá ngoáy hoặc quá bay bướm,… dẫn đến việc nhận dạng không chính xác, tốn nhiều công sức cho việc xác minh số liệu. Nhiều địa phương chưa làm tốt khâu tổng hợp sơ bộ, chưa quan tâm bố trí lực lượng để tự kiểm tra, hoàn thiện phiếu, nên khi gửi báo cáo kết quả sơ bộ và gửi phiếu điều tra về ba Trung tâm Tin học Thống kê vẫn còn nhiều sai sót.

- Một số trường số liệu của phiếu điều tra mẫu đã bị bỏ qua khâu xác minh khối/đống (mass verify) và kiểm tra lôgíc, gây không ít khó khăn trong việc hoàn thiện số liệu phục vụ công bố kết quả điều tra mẫu. Đây là bài học cần rút kinh nghiệm nghiêm túc nhằm bảo đảm chất lượng số liệu điều tra.
- Nhiều máy quét (scanners) đã bị trục trặc nhưng không có thiết bị thay thế kịp thời, việc xử lý sự cố khá chậm, ảnh hưởng đến tiến độ xử lý. Đây là sự cố đột xuất xảy ra ngoài dự kiến, cần rút kinh nghiệm để có phương án thay thế.
- Để khắc phục tình trạng chậm tiến độ do một số trục trặc và sai sót kể trên, khâu nhập tin bằng công nghệ nhận biết ký tự thông minh đã chủ quan không sử dụng các số liệu kiểm tra để chạy ra các cảnh báo hoặc bỏ qua những cảnh báo đó nên số liệu đã không được xem xét xác minh lại kịp thời, làm cho khâu kiểm tra kết quả đầu ra phải làm đi làm lại nhiều lần, lãng phí thời gian và công sức.

Để phát huy hết hiệu quả của công nghệ nhận biết ký tự thông minh, nâng cao chất lượng và giảm thời gian xử lý cho những cuộc điều tra tiếp theo, sau khi kết thúc xử lý Tổng điều tra 2009, cần tổng kết toàn bộ quá trình xử lý để rút kinh nghiệm áp dụng công nghệ này trong tương lai.

4.2 Xử lý số liệu

Nhằm cung cấp kịp thời các kết quả sơ bộ và thông tin cơ bản của Tổng điều tra 2009, Ban Chỉ đạo Tổng điều tra Trung ương đã biên soạn và phát hành báo cáo đầu tiên vào tháng 8 năm 2009. Kết quả sơ bộ của Tổng điều tra phục vụ kịp thời cho công tác đánh giá tình hình thực hiện các kế hoạch và chính sách phát triển kinh tế - xã hội các năm 2001 đến năm 2010, xây dựng kế hoạch và chính sách phát triển kinh tế - xã hội thời kỳ từ năm 2011 đến năm 2020.

Tiếp sau đó kết quả điều tra mẫu được công bố vào tháng 12 năm 2009 với hai ấn phẩm là cuốn sách bỏ túi " Các kết quả suy rộng mẫu", và báo cáo phân tích "Các kết quả chủ yếu".

Kết quả toàn bộ được xuất bản và phổ biến vào tháng 7 năm 2010 bằng hai ấn phẩm sách bỏ túi " Một số chỉ tiêu chủ yếu" và "Kết quả toàn bộ". Đồng thời kết quả toàn bộ dưới dạng biểu tổng hợp bằng file excel cũng được cung cấp cho các tỉnh/thành phố, Bộ/Ngành, và người dùng tin trong và ngoài nước. Người sử dụng đã đánh giá cao kết quả của Tổng điều tra và cho rằng cuộc Tổng điều tra đã cung cấp nguồn thông tin đồ sộ và có chất lượng đáp ứng được mục tiêu đề ra và thoả mãn nhu cầu của những người dùng tin chính.

Ngay sau khi kết thúc điều tra tại thực địa, Tổng cục Thống kê đã phối hợp với UNFPA tổ chức các cuộc hội thảo chuẩn bị cho phân tích sâu các chủ đề của Tổng điều tra 2009. Nhờ vậy, các chuyên khảo của Tổng điều tra thu hút được sự quan tâm của xã hội.

Sau đây là các sản phẩm chính khác của Tổng điều tra 2009 được sản xuất và phát hành như sau:

Tên sản phẩm	Thời gian phát hành
1. Đĩa CD ROM kết quả mẫu và số liệu micro mẫu (kèm theo phần mềm lập biểu) của tỉnh/thành phố	Tháng 12 năm 2010
2. Đĩa CD ROM kết quả mẫu và số liệu micro mẫu (kèm theo phần mềm lập biểu) của cả nước	Tháng 12 năm 2010

Tên sản phẩm	Thời gian phát hành
3. Đĩa CD ROM kết quả toàn bộ và số liệu micro toàn bộ (kèm theo phần mềm lập biểu) của tỉnh/thành phố	Tháng 12 năm 2010
4. Đĩa CD ROM kết quả toàn bộ của của cả nước	Tháng 12 năm 2010
5. Ấn phẩm "Dự báo dân số Việt Nam 2009-2049	Tháng 2 năm 2011
6. Chuyên khảo "Cấu trúc tuổi - giới tính và tình trạng hôn nhân của dân số Việt Nam	Tháng 4 năm 2011
7. Chuyên khảo "Phân tích các chỉ số chủ yếu về giáo dục của Việt nam	Tháng 5 năm 2011
8. Chuyên khảo " Di cư và đô thị hoá ở Việt Nam: thực trạng, xu hướng và những khác biệt"	Tháng 5 năm 2011
9. Chuyên khảo "Tỷ số giới tính khi sinh ở Việt Nam: Các bằng chứng mới về thực trạng, xu hướng và những khác biệt"	Tháng 5 năm 2011
10. Mức sinh và mức chết ở Việt Nam: Thực trạng, xu hướng và những khác biệt"	Tháng 11 năm 2011
11. Báo cáo hành chính của Tổng điều tra dân số và nhà ở 2009, 2 Tập	Tháng 11 năm 2011
12. Kho dữ liệu Tổng điều tra dân số và nhà ở 2009 trên web và Atlas của Tổng điều tra	Tháng 11 năm 2011