

Rus.

## MỘT SỐ PHƯƠNG PHÁP TỰ ĐỘNG HÓA TỔNG HỢP VĂN BẢN

ĐÓ VIỆT-NGA

Tự động hóa trong lĩnh vực hành chính sự nghiệp ngày càng trở nên vấn đề cấp thiết của một nền kinh tế hiện đại. Tốc độ cơ khí hóa nhanh chóng trong mọi lĩnh vực kinh tế quốc dân gây ra khoảng cách lớn giữa sản xuất và bộ máy điều hành. Sự chênh lệch giữa một bên là nền sản xuất được tự động hóa cao và một bên là bộ máy hành chính công kênh, lâu đời, ít thừa theo kiểu thủ công làm giảm đáng kể năng suất lao động. Theo thống kê trong khi phần lớn nền kinh tế ở một số nước phương Tây cho thấy ở nhiều lĩnh vực, trên 50% giá thành sản phẩm là do chi phí cho các giấy tờ, các khâu hành chính để sản phẩm có thể đi từ nguyên liệu cho đến tay người tiêu dùng [1].

Do đó việc tăng năng suất lao động trong các khâu hành chính ngày càng được quan tâm ở các nước tiên tiến. Trong công việc hành chính, các phần việc liên quan tới văn bản, giấy tờ lại là phần việc phổ biến và rắc rối nhất. Vì vậy, trong công cuộc "Tự động hóa văn phòng" việc đầu tiên người ta quan tâm tới là tự động hóa các khâu văn bản giấy tờ.

Hàng loạt các kiểu hệ thống soạn thảo văn bản được ra đời từ thập kỉ 70. Ban đầu chúng được dùng để phục vụ cho các thảo chương viên, giảm đáng kể lao động của họ. Dần dần các hệ thống này được cải tiến và được dùng để phục vụ cả việc soạn thảo văn bản. Ngày nay trên thế giới đã có tới hàng trăm hệ thống soạn thảo khác nhau, ngày càng thích hợp hơn với nhu cầu của người sử dụng.

Các hệ thống soạn thảo văn bản hiện hành có thể chia làm hai kiểu:

1. Kiểu vận năng: Có thể sử dụng chúng để soạn thảo một văn bản bất kỳ (ví dụ như hệ thống kiểu WORDSTAR). Các hệ thống này có ưu thế ở điểm vận năng, nhưng thường bị hạn chế ở tốc độ (tốc độ vào dữ liệu ngang với tốc độ đánh máy chữ).

2. Kiểu điền vào mẫu có sẵn: Kiểu này thường được dùng trong các công đoạn sản xuất chuyên dụng. Các văn bản được thảo sẵn để chừa cho một vài thông số, và hàng ngày, khi cần thiết các chỗ trống được điền thích hợp. Cách soạn thảo này cho tốc độ lớn, nhưng lại bị hạn chế bởi khuôn khổ của văn bản thảo sẵn. Mặt khác không phải loại văn bản nào cũng có thể mẫu hóa theo cách đó.

Trong bài báo này tác giả muốn trình bày một cách tiếp cận dung hòa giữa hai cách trên nhằm xây dựng các phương pháp tổng hợp văn bản cho phép tăng tốc độ soạn thảo văn bản, mặt khác không hạn chế tính vận năng của hệ thống.

Tác giả đề xuất hai phương pháp tổng hợp văn bản và cách thiết kế trên máy tính điện tử.

Trong bài có giới thiệu hệ thống tự động hóa soạn thảo văn bản đã được xây dựng dựa trên cơ sở phương pháp đã đề xuất.

### I - CÁCH TIẾP CẬN NHẪM XÂY DỰNG PHƯƠNG PHÁP TỔNG HỢP VĂN BẢN

Theo cách tiếp cận này một hệ thống tổng hợp văn bản được xây dựng từ hai thành phần: phần hệ thống cơ sở và phần đặc thù.

Phần hệ thống cơ sở làm công việc điều hành, tổng hợp văn bản dựa trên thông tin lấy từ phần đặc thù.

Phần đặc thù chứa đựng các thông tin số đo của một loại văn bản.

Phần hệ thống được xây dựng bởi người lập trình, còn phần đặc thù do chuyên gia, chuyên trách việc soạn thảo tài liệu tự tạo lập trên cơ sở phân tích và kinh nghiệm sẵn có về đặc thù của loại tài liệu trong phạm vi trách nhiệm của mình.

Mọi hệ thống gồm phần hệ thống cơ sở và phần đặc thù tạo nên một hệ chuyên đề soạn thảo một vài loại thư tín, công văn, tài liệu.

Bằng cách đó, hệ thống có thể tăng tốc độ soạn thảo một cách đáng kể dựa trên các thông tin đặc thù đã được phân tích và lập sẵn. Mặt khác, tính vận năng của hệ thống có thể được tăng thêm bằng cách mở rộng phần đặc thù.

Như vậy, ta có thể thu được một hệ thống có tốc độ soạn thảo nhanh và theo một nghĩa nào đó, vẫn có tính vận năng. Hơn thế nữa, hệ thống luôn có thể được bổ sung, mở rộng bởi công lao động của bản thân người sử dụng, thường là các nhân viên thư ký.

Dựa trên cách tiếp cận đã trình bày, tác giả đề xuất hai phương pháp đồng hợp văn bản.

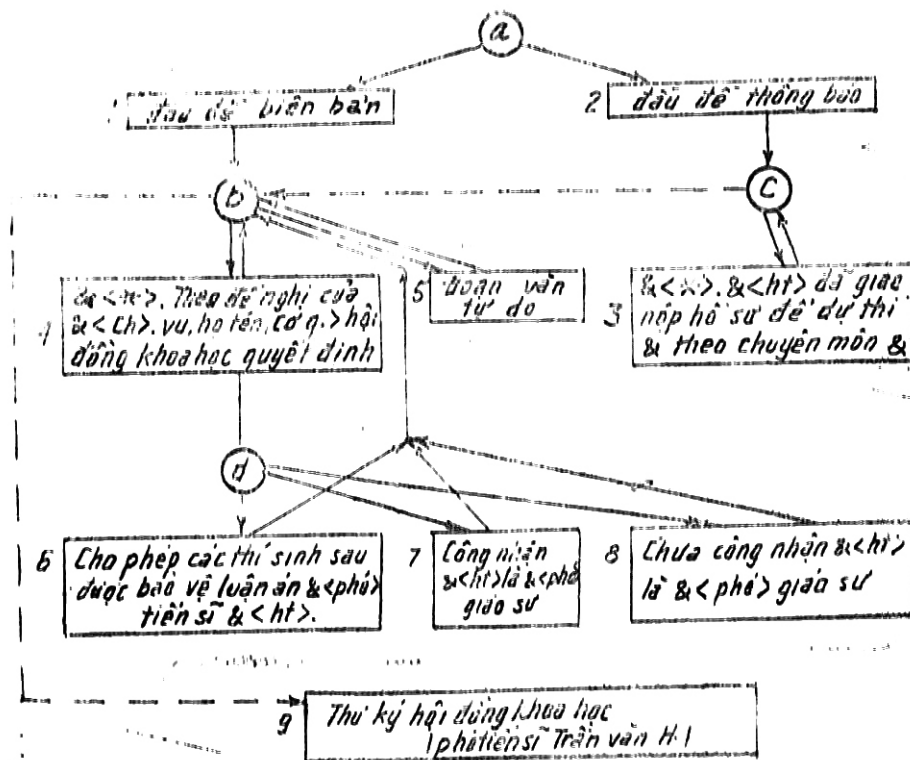
## II - PHƯƠNG PHÁP TỔNG HỢP VĂN BẢN DỰA TRÊN VĂN BẢN MẪU

Phương pháp này giả thiết rằng, các văn bản được tạo ra từ các đoạn văn. Các đoạn văn lại có thể chuẩn hóa, thảo sẵn, lưu trữ trong máy tính và được sử dụng trong quá trình soạn thảo. Các đoạn văn soạn sẵn cũng có thể chứa các phần trống được điền giá trị cụ thể trong quá trình soạn thảo.

Phần đặc thù ở phương pháp này chứa một loạt các maket (văn bản mẫu). Mọi văn bản mẫu chứa một số các đoạn văn mẫu, các phần trống để điền và các thứ tự xuất hiện cho phép của các đoạn văn.

Phần hệ thống cơ sở ở phương pháp này là một chương trình cho phép sử dụng maket để soạn thảo một hoặc một loạt văn bản, trong đó có thể lựa chọn đoạn văn, lập một đoạn văn ở nhiều chỗ, điền giá trị phần trống từ bàn phím hoặc từ một danh sách soạn sẵn.

Ta có thể xét phương pháp này trên một ví dụ. Với mẫu văn bản trên hình 1 ta có thể xây dựng các biên bản và các thông báo khác nhau. Một trong các biên bản có ở hình 2.



Hình 1

## BIÊN BẢN SỐ 16

Phiên họp Hội đồng Khoa học viện...

ngày 12-04-1986

Cuộc họp bắt đầu từ 14h đến 16h tại hội trường của Hội đồng Khoa học viện...

Trong cuộc họp có sự tham gia của các vị: giáo sư tiến sĩ Lê Thanh M., giáo sư p/s Trần Việt A....

- 4  
+  
6 1. Theo đề nghị của viện KHTTĐK Hội đồng Khoa học quyết định cho các thí sinh sau đây bảo vệ luận án phó tiến sĩ: Nguyễn Văn A và Lê thị B.
- 4  
+  
7 2. Theo đề nghị của chủ tịch Hội đồng giám khảo chọn giáo sư. Hội đồng khoa học quyết định công nhận p/s Trần H là phó giáo sư.
- 5 3. Hội đồng khoa học quyết định trao tặng bằng khen kèm phần thưởng hiện vật trị giá 1000đ cho...

9

*Thư ký Hội đồng khoa học  
p/s Trần Văn H*

### Hình 2

Trong hình 1 các mũi tên liền nét chỉ thứ tự xuất hiện của các đoạn văn, các mũi tên không liền nét chỉ điểm ra của các vòng lặp. Các biên bản, thông báo lập từ mẫu văn bản này là một đường đi (có thể có vòng lặp) từ điểm a cho tới điểm 9. Các nút tròn a, b, c... chỉ các nút cho phép lựa chọn. Biên bản trên hình 2 được tạo lập khi hệ thống đi từ a tới 9 theo đường a1b4d6b4d7b5b9. Các chỗ được điền thêm từ bản phim được đánh dấu trong ví dụ bằng cách gạch dưới chân.

Cần lưu ý là ở đây thứ tự xuất hiện của các đoạn văn dường như gõ văn bản theo một khuôn mẫu, nhưng thực ra thứ tự này hoàn toàn do người sử dụng quyết định, thậm chí có thể vứt bỏ thứ tự (ví dụ, các đoạn văn 4,5 hoặc 6, 7, 8 là không có thứ tự trước sau. Giữa chúng có thể lựa chọn tùy ý). Thứ tự giữa các đoạn văn được dựa vào văn bản mẫu với một dụng ý hoàn toàn khác. Ta xét, giả sử trong văn bản mẫu có phần về việc bầu cử. Đoạn văn đầu tiên về phần bầu cử phải là đoạn văn giới thiệu mục đích, sau đó đến thành phần bầu cử và cuối cùng là kết quả. Ba đoạn văn này nhất thiết phải theo thứ tự đã mô tả và nếu đoạn văn đầu tiên không xuất hiện thì cả hai đoạn văn sau cũng bị bỏ qua, có nghĩa là trong văn bản không có phần bầu cử. Qua ví dụ ta có thể nhận thấy là thứ tự được đưa vào để giúp người sử dụng có thể lựa chọn theo nội dung các đoạn văn.

### III - HỆ THỐNG TỰ ĐỘNG HÓA SOẠN THEO VĂN BẢN (АСИДО)

Hệ thống được thiết kế và cài đặt dựa trên phương pháp đã được mô tả ở phần II.

Hệ thống cơ sở bao gồm ba modul. Modul chính thực hiện việc hiện các đoạn văn từ văn bản mẫu ra màn hình, cho phép người sử dụng được lựa chọn đoạn văn cần thiết điền các chỗ trống hoặc đưa đoạn văn mới vào văn bản từ bản phim. Quá trình soạn thảo chỉ kết thúc khi văn bản được hoàn thành.

Hệ thống có cả các chức năng kể từ bên phải, cho phép phân chia từ đối với các ngôn ngữ Anh, Nga, Bun; chức năng phân trung, soạn thảo một loạt thư tín cùng kiểu cho nhiều người nhận khác nhau, v.v...

Hệ thống đã được thử nghiệm cho các công văn, biên bản, tài liệu hành chính của Hội đồng Khoa học, viện Hàm lâm Bungari. Hệ cũng đã được thử nghiệm trong việc soạn thảo văn bản tiếng Việt.

Hệ thống АСИДО được viết trên ngôn ngữ lập trình PASCAL cho máy vi tính 16 bit kiểu PC/IBM hoặc tương thích, với cấu hình tối thiểu.

#### IV — PHƯƠNG PHÁP TỰ ĐỘNG HÓA TỔNG HỢP VĂN BẢN BẰNG CÁCH VIẾT TẮT

Phương pháp này giả thiết là trong mọi lĩnh vực đời sống hàng ngày các từ ngữ được dùng thường có đặc thù riêng. Có nghĩa là một số từ hoặc nhóm từ được dùng lặp đi lặp lại nhiều lần. Sử dụng tính chất này ta có thể tạo lập một hệ thống cho phép ghi tóm tắt một số từ, nhóm từ.

Phần đặc thù trong phương pháp này là một từ điển do bản thân người sử dụng lập ra.

Phần hệ thống cơ sở là một bộ soạn thảo văn bản có khả năng hiểu được cách viết tắt của người sử dụng.

Để hiểu rõ hơn, ta lấy một ví dụ viết tắt thường gặp trong văn viết « v.v... » có nghĩa là « vân vân » hay « đ/c » có nghĩa là « đồng chí ». Trong hệ thống cho phép cả hai kiểu viết tắt: viết tắt bằng nhóm chữ đầu tiên và dấu viết tắt (ví dụ như « v. ») hoặc viết tắt bằng kí hiệu riêng. Để viết tắt theo kí hiệu riêng người sử dụng cần phải khai báo trước mình kí hiệu của mình khi đưa từ tương ứng vào từ điển. Khi viết tắt theo nhóm chữ đầu người sử dụng có thể viết dài tùy ý (ví dụ có thể viết « v. », « vâ. », « vân. » thay cho từ « vân » có trong từ điển).

Phương pháp đặc biệt hiệu quả đối với các ngôn ngữ đa âm với các từ dài.

Hai phương pháp trên ngoài việc tăng đáng kể tốc độ soạn thảo dữ liệu còn cho phép giảm phần lớn các lỗi chính tả, cú pháp, lỗi hành văn. Các phương pháp này đòi hỏi phải được thử nghiệm, nghiên cứu sâu sắc cách làm việc của các nhân viên thư ký. Hiện nay các phương pháp này vẫn đang tiếp tục được nghiên cứu và phát triển.

Nhận ngày 1-3-87

#### TÀI LIỆU THAM KHẢO

1. К. Кочетков, Автоматизация конторского труда в США издательство наука, 1985.
2. П. Бърнев, До Виет Ига, Система автоматизированного составления документов. Сб. Математика и математическо образование, 1986.
3. До Виет Ига, Система автоматизированного составления свободных текстов. Труды конференции КИВЗТ, Будапешт, 1986.
4. До Виет Ига, Система автоматизированного синтеза документов. Труды школы «Программирование 86», 1986.

#### РЕЗЮМЕ

#### НЕКОТОРЫЕ МЕТОДЫ АВТОМАТИЗИРОВАННОГО СИНТЕЗА ДОКУМЕНТОВ

В работе описываются два метода автоматизированного синтеза документов: метод синтеза по текстовым образцам и метод синтеза по сокращенным словам. Первый метод служит основой для построения автоматизированной системы составления документов АСИДО. Данная система экспериментировалась в некоторых учреждениях. Система может обрабатывать тексты написанные на русском, болгарском и Вьетнамском языках.