

## MỘT PHƯƠNG PHÁP XỬ LÝ IN CHỮ VIỆT VÀ CHỮ NGA CHO CÁC MÁY IN KIỂU EPSON--FX85, FX100, FX1000

TRẦN BÁ THÁI--PHẠM MẠNH CỒN

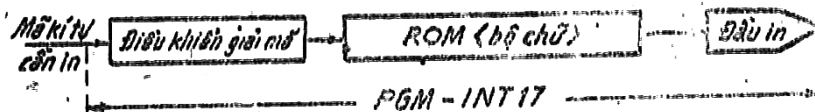
Gần đây, với việc thâm nhập hàng loạt máy vi tính IBM-PC và với sự xuất hiện các chương trình xử lý văn bản như Edit, Editin hay Wordperfect... trên thị trường Việt Nam thì việc sử dụng chữ Việt cũng như đòi hỏi thể hiện ra chữ Việt trên máy in là một hướng được nhiều cơ quan nghiên cứu quan tâm.

Bài viết này chỉ giới thiệu một phần nhỏ của việc cài đặt chữ Việt lên máy vi tính IBM-PC, đó là phương pháp xử lý in chữ Việt cho các máy in kiểu EPSON-FX. Đây cũng là một trong những kết quả của nhóm nghiên cứu đã tài cấp nhà nước « máy vi tính 16-32 bit (18A-03-01) » do phòng KTS chủ trì.

Việc cài đặt chữ Nga và việc in chữ Nga trên máy EPSON-FX chỉ là một hệ quả của việc cài đặt chữ Việt lên hệ vi xử lý.

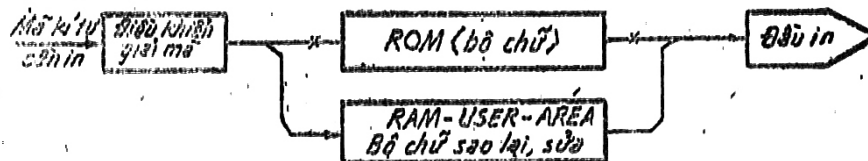
### 1. Khả năng điều khiển ký tự in của máy in EPSON-FX.

Ở chế độ làm việc bình thường, mỗi khi nhận được một ký tự từ bàn phím, đầu in sẽ in sau khi tìm được ký tự đó nằm ở đâu trong bộ chữ của máy in. Quá trình này được minh họa như sau:



Làm tất cả công việc này là nhiệm vụ của một chương trình mà địa chỉ của nó nằm ở nội dung của vectơ ngắt INT 17 (Interruption) của hệ thống. Ta gọi nó là PGM-INT-17.

Trong máy in có một vùng nhớ RAM, trên đó ta có thể sao lại nguyên si bộ chữ từ ROM và sau đó định nghĩa lại ký tự trên vùng này. Cơ chế này được minh họa sau đây:



Thứ tự của quá trình trên như sau (các lệnh điều khiển máy in viết kèm):

- Nạp bộ chữ từ ROM sang RAM:  
ESC: (1B 3A 0 0 0).
- Định nghĩa lại các ký tự:  
ESC & (1B 26 0 d0 d1 d2... data...).
- Xác định chế độ làm việc từ vùng RAM:  
ESC % (1B 25 n

n = 1: in ký tự từ ROM  
n = 0: in ký tự từ RAM).

**2. Cơ chế in chữ Việt trên các máy in EPSON-FX trong tổ hợp bàn phím, màn hình tiếng Việt**

**a) Định nghĩa bộ chữ Việt cho máy in EPSON-FX**

Sẽ rất tiện cho người sử dụng khi ta có thể dùng tiếng Việt và dùng cả tiếng Anh (hoặc Pháp) để xử lý văn bản. Như thế, toàn bộ các ký tự theo chuẩn ASCII từ 20H đến 79H tương ứng với các chữ cái từ a đến z và các dấu, các ký hiệu đều được nạp từ bộ chữ ROM sang RAM. Sẽ không có sự thay đổi nào đối với các ký tự có mã nằm trong vùng địa chỉ này. Ta sẽ sử dụng các ký tự có mã ASCII từ 80H trở lên để định nghĩa các ký tự riêng biệt của chữ Việt.

Với mỗi lệnh ESC & ta có một hích chữ nhật 8 hàng và 12 cột để định nghĩa một ký tự, kể cả 2 dòng dưới cùng cho các chữ có đuôi như y, g. Trong một diện tích hẹp như vậy ta không thể định nghĩa một nguyên âm của tiếng Việt cộng với một trong các dấu thanh hoặc dấu âm, càng không thể cấu tạo được một nguyên âm cộng với dấu âm và dấu thanh. Do vậy, ta sẽ định nghĩa tất cả các dấu âm và dấu thanh, các dấu âm cộng dấu thanh một cách riêng biệt.

Các dấu này là:

^ (huyền); ? (hỏi); ~ (ngã); / (sắc)  
 ^ ; ^ ; ^ ; ^  
 ^ ; ^ ; ^ ; ^

Loại trừ các ký tự: a, o, e, i, o, u mà ta cũng sẽ định nghĩa riêng trong vùng địa chỉ lớn hơn 80H. Như thế, muốn in chữ á thì đầu in sẽ in 2 lần; lần 1 - in dấu mũ, lần hai - in chữ a.

Việc xử lý cách hàng giữa chữ và dấu hoàn toàn thực hiện được bằng các lệnh điều khiển máy in mà ta giữ ở từ bàn phím.

Một bản định nghĩa cho các ký tự dấu theo chuẩn do hội nghị về "xử lý văn bản tiếng Việt trên các máy vi tính" qui định chiếm 1000 byte.

**b) Cách in một văn bản tiếng Việt**

Giả sử cần in một câu có nội dung sau đây:

« Nhà nước xã hội chủ nghĩa »

Như đã trình bày ở trên, một chữ Việt được chia làm 2 phần, phần dấu và phần chữ. Các phần này sẽ được xếp vào hai bộ đệm: bộ đệm dấu (BD) và bộ đệm chữ (BC). Hai bộ đệm này có độ dài bằng nhau và đều nằm trong bộ nhớ của hệ.

Trước hết, ta xếp dấu của ký tự đầu vào BD, sau đó xếp chính ký tự đó vào BC. Công việc này được lặp đi lặp lại cho đến ký tự cuối cùng của câu.

Ở các vị trí của BD tương ứng với ký tự chữ không dấu của BC ta điền vào đó dấu cách (space-blanc).

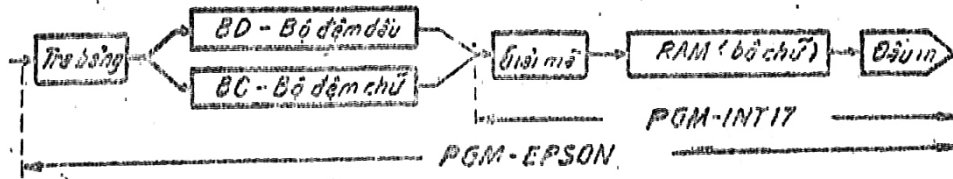
Có sơ đồ sau:

BD															
Bộ đệm dấu															
BC	n	h	a	n	u	o	c	x	a	h	o	i	c	h	u
Bộ đệm chữ															

Công việc còn lại là nhiệm vụ của PGM-IST-17, từng ký tự một, in hết cả nội dung của BD, xuống hàng và cũng từng ký tự một, in hết nội dung của BC. Cũng phải kể đến việc xử lý cách hàng giữa hàng dấu và hàng chữ trong khi in.

Làm tất cả công việc đó là nhiệm vụ của một chương trình, được gọi là chương trình xử lý in PGM-EPSON.

Ta có sơ đồ sau :



Đối với hệ điều hành, mỗi khi ngắt INT 17 được tác dụng, chương trình hệ thống nhảy đến chương trình 'PGM-INT-17' mà địa chỉ của nó được đặt ở nội dung của vectơ ngắt INT17—Đó là chương trình xử lý in một ký tự cho máy in. Như đã trình bày ở trên, trước khi gửi đi in, ký tự cần in được xử lý, phân loại, xếp vào các bộ đệm BD và BC. Chương trình PGM-INT-17 chỉ thực sự được gọi đến khi tổ chức bộ đệm đã đầy. Do đó, mỗi khi vectơ ngắt INT 17 được gọi, thay vì chương trình hệ thống nhảy đến địa chỉ của PGM-INT-17 sẽ nhảy đến địa chỉ của PGM-EPSON. Điều này có nghĩa là ở nội dung của vectơ ngắt INT17, ta đặt địa chỉ của chương trình xử lý in EPSON Chính PGM-EPSON sẽ gọi PGM-INT-17 như một chương trình con (SBR). Hoặc nếu ta đặt địa chỉ của PGM-INT-17 vào địa chỉ của một vectơ còn được phép sử dụng, thì dụ INT 34, thì INT 17 sẽ gọi INT 34.

Cần phải chú ý là: chương trình xử lý in PGM-EPSON là nhiệm vụ phân loại các ký tự in do đó chính PGM-EPSON phải nhận biết được tất cả mã của các ký tự đã qui định cho bàn phím màn hình. Muốn làm điều đó, chương trình PGM-EPSON được trang bị một thủ tục tra « từ điển ». « Từ điển » của PGM-EPSON gồm 3 cột :

cột 1	cột 2	cột 3	chữ Việt
mã ký tự HEX	mã của chữ	mã của dấu	
CO	41	AO	à
C1	41	A1	á

Một « từ điển » như vậy chiếm 072 byte trong chương trình xử lý PGM-EPSON.

c) Lưu trữ của chương trình xử lý in PGM-EPSON

Quá trình in chữ Việt gồm 2 phần :

#### Phần 1

- Định nghĩa bộ chữ cho máy in
  - Lệnh cho máy in làm việc với cùng RAM
  - Thay đổi địa chỉ, đặt địa chỉ của PGM-EPSON vào nội dung của INT17
- Đặt PGM-INT 17 vào một ngắt (INT 34)

#### Phần 2

- Phân loại ký tự
- Tạo hai bộ đệm BC, BD.
- Gọi PGM-INT-17 để in các ký tự từ BD và BC.

Các công việc của phần 2 được lặp đi lặp lại cho đến khi in xong văn bản.

### 3. Cách in văn bản trong tiếng Nga

Đối với văn bản tiếng Nga, việc xử lý in đơn giản hơn rất nhiều so với xử lý in trong tiếng Việt.

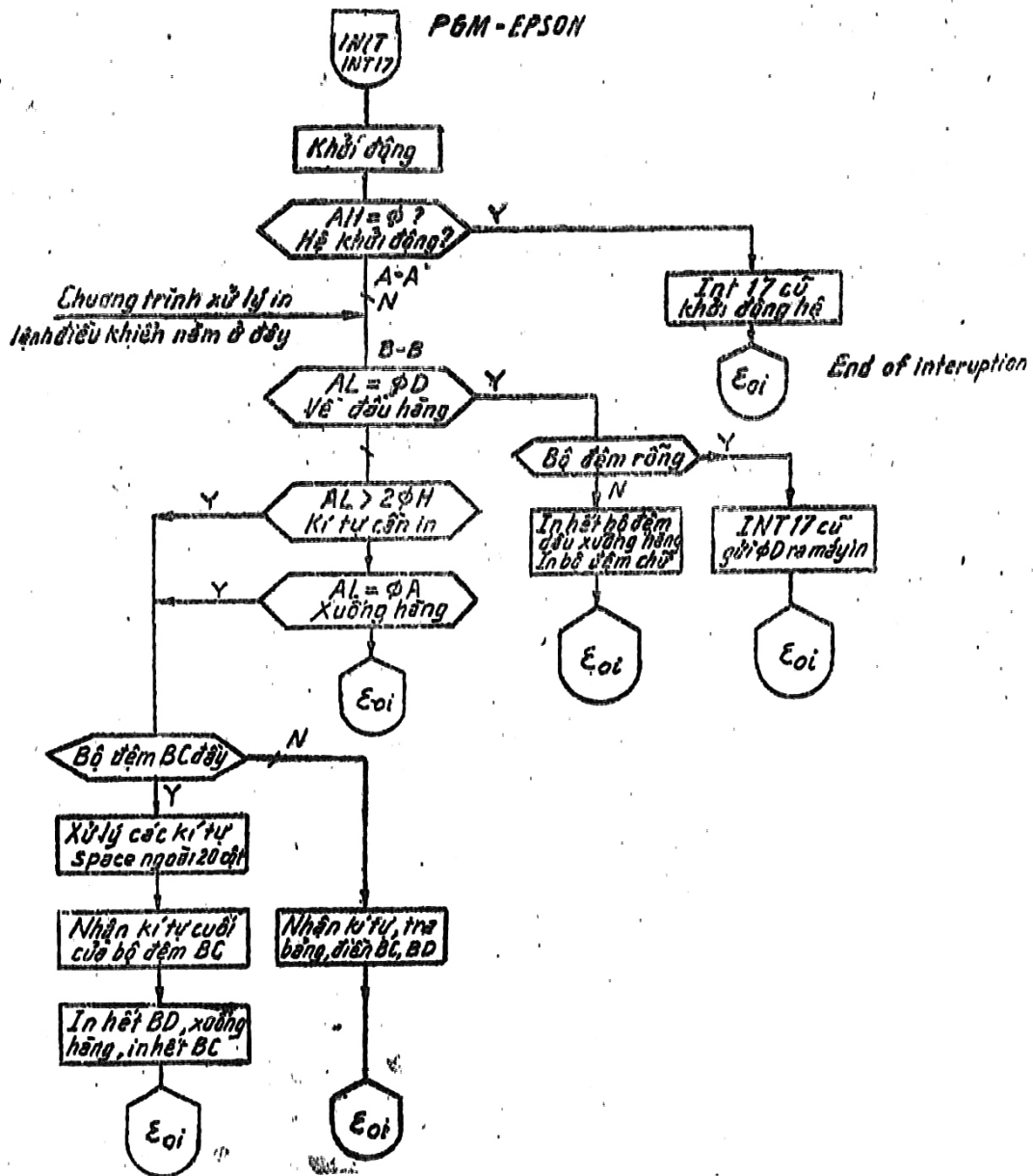
Tất cả công việc xử lý in trong tiếng Nga chỉ nằm gọn trong phần 1 của xử lý in trong tiếng Việt.

- Định nghĩa bộ chữ cho máy in.

- Lệnh cho máy in làm việc với vùng RAM. Vì trong tiếng Nga không có ký tự dấu, nên ta không phải quan tâm đến khái niệm về tổ chức bộ đệm. Mỗi khi INT 17 được gọi thì chương trình hệ thống chạy trực tiếp chương trình PGM-INT 17, lần lượt tăng ký tự in hết văn bản cần in.

Để định nghĩa bộ chữ cho tiếng Nga, ta cần một vùng nhớ có dung lượng 1152 byte.

#### 4. Cách xử lý các lệnh điều khiển máy in EPSON-FX.



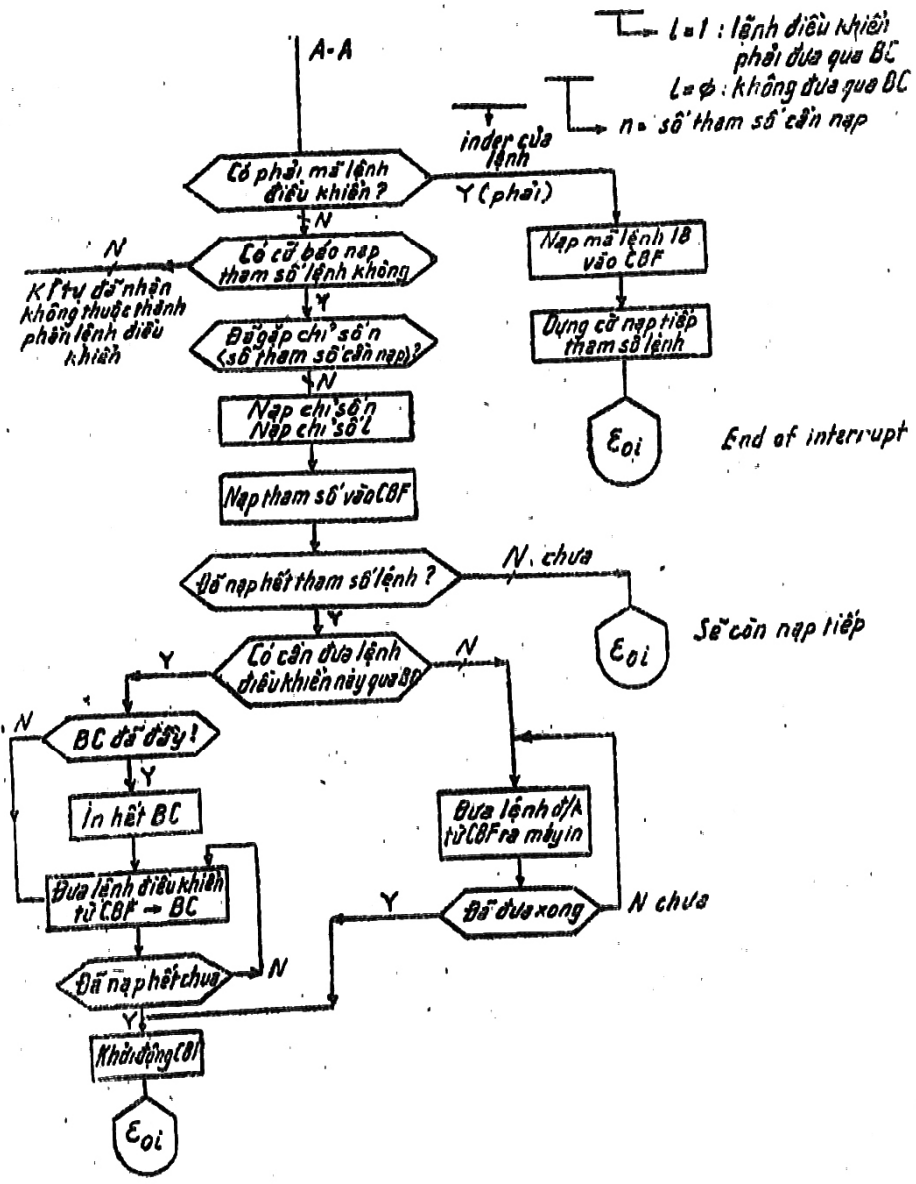
Lưu đồ chương trình PGM-EPSON

Các lệnh điều khiển gồm có mã lệnh và một số lượng khác nhau các thuộc tính (attribut). Do đó, chương trình xử lý in phải biết được khi nào chúng bắt đầu và khi nào thì chúng kết thúc? Cũng phải xét đến khả năng xem lệnh điều khiển đó có phải đưa qua bộ đệm hay được đưa trực tiếp ra máy in?

Ta sẽ dùng một bộ đệm, tại đây lệnh điều khiển và các thuộc tính của nó sẽ được nạp vào trước khi chúng được đưa qua bộ đệm ký tự chữ hay đưa trực tiếp ra máy in.

Lưu đồ chương trình xử lý lệnh điều khiển:

1000	1001	1002	1003						
CBF	φ	φ	n	l					lφ



Nhận ngày 19-11-1987