

9/9.2

Filtrage de fichiers ASCII

Si vous travaillez avec un traitement de texte évolué sur votre CPC (*Pocket Wordstar* par exemple), vous connaissez très certainement une de ses fonctions les plus évoluées : « Recherche et Remplace ». Cette fonction recherche une chaîne alphanumérique dans le texte en mémoire et la remplace par une autre chaîne alphanumérique.

Si les modifications à apporter aux fichiers texte que vous utilisez sont nombreuses et/ou systématiques, vous préférerez certainement passer par le programme utilitaire que nous présentons ici. Ce programme permet de rechercher et de remplacer automatiquement un nombre quelconque de chaînes alphanumériques. La seule limite est la taille de mémoire RAM disponible.

Afin d'automatiser les remplacements, vous devez dans un premier temps créer un fichier de directives qui contient le libellé des chaînes à remplacer et le libellé des chaînes qui les remplacent.

Pour ce faire, il vous suffit de connaître :

- le nombre de chaînes à remplacer,
- leur libellé et celui des chaînes qui les remplacent, et de donner un nom au fichier dans lequel ces données vont être stockées.

Ces données apparaissent dans le programme Basic ci-dessous :

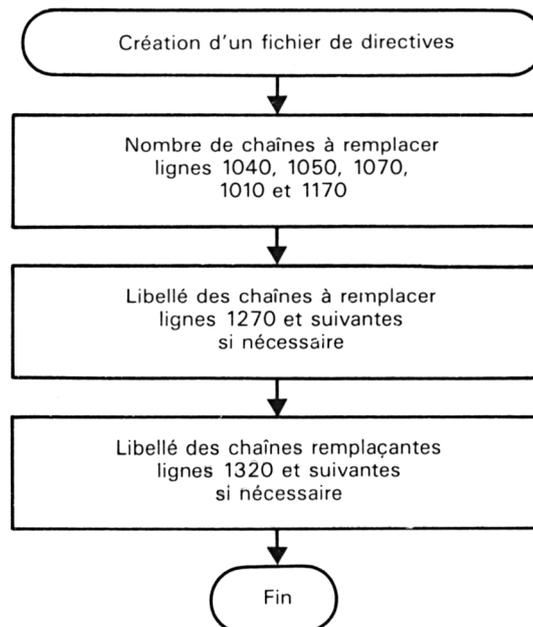
```

1000 '=====
1010 '   CONSTITUTION D'UN FICHER DE DIRECTIVES
1020 '=====
1030 '
1040 DIM o$(100) 'Chaines recherchees
1050 DIM d$(100) 'Chaines rempalcantes
1060 '
1070 FOR i=1 TO 10
1080   READ o$(i) 'Lecture des chaines a remplacer
1090 NEXT i
1100 '
1110 FOR i=1 TO 10
1120   READ d$(i) 'Lecture des chaines rempalcantes

```

```
1130 NEXT i
1140 '
1150 fr$="regle.dat" 'Nom du fichier de regles
1160 OPENOUT fr$
1170 FOR i=1 TO 10
1180   PRINT #9,o$(i) 'Sauvegarde des chaines a remplacer
1190   PRINT #9,d$(i) 'Sauvegarde des chaines rempalcantes
1200 NEXT i
1210 CLOSEOUT
1220 '
1230 '-----
1240 'Chaines a remplacer
1250 '-----
1260 '
1270 DATA a,b,c,de,f,g,h,1,2,30
1280 '-----
1290 'Chaines rempalcantes
1300 '-----
1310 '
1320 DATA A,B,C,De,F,G,H,un,deux,trente
```

L'ordinogramme suivant vous indique le numéro de lignes où reporter les données.

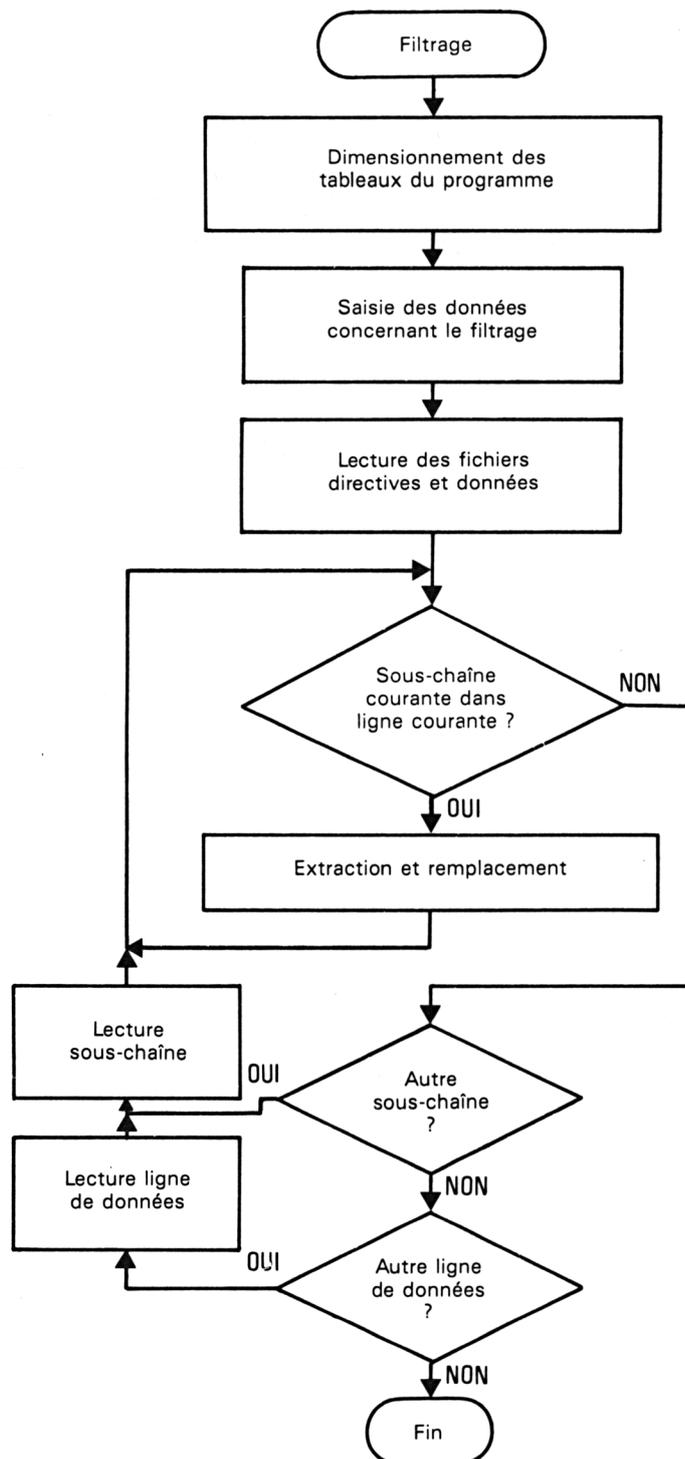


Cette première étape achevée, il vous suffit d'exécuter le programme de filtrage, et d'entrer :

- le nom du fichier de directives ;
- le nom du fichier à filtrer ;
- le nom du fichier résultant.

Quelques instants après avoir entré ces informations, le message « **Filtrage en cours** » apparaît sur l'écran. Ce message est suivi de l'affichage d'un ou de plusieurs points décimaux. Chaque point correspond au traitement d'une ligne dans le fichier à filtrer. Lorsque le filtrage est terminé, le message « **Filtrage terminé** » est affiché sur l'écran.

La logique du programme de filtrage est la suivante :



Le listing du programme de filtrage est le suivant :

```

1000 '=====
1010 ' FILTRAGE DE FICHIERS ASCII
1020 '=====
1030 '
1040 '-----
1050 'Programme principal
1060 '-----
1070 '
1080 GOSUB 2000 'Initialisation
1090 GOSUB 3000 'Entree des donnees
1100 GOSUB 4000 'Conversion
1110 END
2000 '-----
2010 ' Initialisation
2020 '-----
2030 '
2040 DIM o$(100) 'Chaines origine
2050 DIM d$(100) 'Chaines destination
2060 DIM f$(100) 'lignes a convertir
2070 '
2080 RETURN
3000 '-----
3010 'Entree des donnees
3020 '-----
3030 '
3040 ON ERROR GOTO 10000
3050 MODE 2
3060 PRINT"Filtrage de fichiers ASCII"
3070 PRINT"-----"
3080 PRINT:PRINT:PRINT
3090 INPUT"Nom du fichiers de directives : ";fd$
3100 a=1:OPENIN fd$:CLOSEIN
3110 INPUT "Nom du fichier a filtrer : ";ff$
3120 a=2:OPENIN ff$:CLOSEIN
3130 INPUT "Nom du fichier resultat : ";fr$
3140 RETURN
4000 '-----
4010 ' Conversion
4020 '-----
4030 '
4040 a=3
4050 OPENIN fd$
4060 i=1
4070 WHILE NOT EOF
4080 INPUT #9,o$(i) 'Chaines a convertir
4090 INPUT #9,d$(i) 'Conversions
4100 i=i+1
4110 WEND
4120 Nbre=i-1 'Nombre de regles de conversion
4130 CLOSEIN
4140 '
4150 OPENIN ff$
4160 i=1 'Initialisation
4170 WHILE NOT EOF
4180 INPUT #9,f$(i) 'Lignes a filtrer
4190 i=i+1
4200 WEND
4210 nbli=i-1 'Nombre de lignes a traiter

```

Partie 9 : Programmes

```

4220 CLOSEIN
4230 '
4240 '
4250 PRINT:PRINT"Filtrage en cours ..."
4260 OPENOUT fr$ 'Ouverture du fichier resultat
4270 FOR i=1 TO nbli
4280   FOR j=1 TO nbre
4290     bis=1
4300     WHILE bis=1
4310       p=INSTR(f$(i),o$(j))
4320       IF p=0 THEN 4420 'Fin de la recherche
4330       ta#=f$(i) 'Memorisation tampon
4340       lf=LEN(f$(i)) 'Longueur de la chaine origine
4350       ld=LEN(d$(j)) 'Longueur de la chaine remplaceante
4360       lo=LEN(o$(j)) 'Longueur de la chaine a remplacer
4370       f$(i)=" 'Mise a vide de la chaine origine
4380       IF p<>1 THEN f$(i)=MID$(ta#,1,p-1)
4390       f$(i)=f$(i)+d$(j) 'Modification
4400       f$(i)=f$(i)+MID$(ta#,p+lo,lf-p-lo+1) 'Extraction 2eme partie
4410       GOTO 4430 'Suite de la recherche
4420       bis=0
4430     WEND
4440   NEXT j
4450   PRINT #9,f$(i)
4460   PRINT". "; 'pour patienter
4470 NEXT i
4480 CLOSEOUT 'Fermeture du fichier resultat
4490 PRINT:PRINT"Filtrage termine"
4500 RETURN
10000 '-----
10010 ' Routine d'erreur
10020 '-----
10030 PRINT:PRINT"Fichier inexistant"
10040 IF a=1 THEN RESUME 3090
10050 IF a=2 THEN RESUME 3110

```

- Lignes 1000 à 1110 : Programme principal.
- Lignes 2000 à 2080 : Dimensionnement des tableaux.
- Lignes 3000 à 3140 : Saisie des données concernant le filtrage.
- Lignes 4000 à 4500 : Filtrage.
- Lignes 4070 à 4130 : Lecture des chaînes à convertir.
- Lignes 4150 à 4220 : Lecture des chaînes de conversion.
- Lignes 4270 à 4470 : Algorithme de filtrage.
- Lignes 4310 : Recherche d'une chaîne dans la ligne courante.
- Ligne 4380 : Isolement 1^{re} partie de la chaîne.
- Ligne 4390 : Conversion.
- Ligne 4400 : Isolement 2^e partie de la chaîne.
- Lignes 10000 à 10050 : Routine d'erreur activée si le fichier de directives ou le fichier à filtrer n'existent pas sur le disque courant.