TD/TP 6 : Fichiers

Gilles Peskine

15 novembre 2002

1 Taille d'un fichier

Proposer des méthodes pour obtenir la taille d'un fichier. Évaluer leurs mérites relatifs en terme de portabilité, de champ d'application, d'efficacité...

2 Concaténation de fichiers

Écrire un programme qui concatène (met bout à bout) les fichiers qui lui sont passés en arguments. C'est la commande Unix `cat`. Par exemple, `cat toto tutu` affiche sur sa sortie standard le contenu du fichier `toto` puis celui du fichier `tutu`. Ne pas oublier de traiter le cas où un fichier n'existe pas.

3 Recherche d'une sous-chaîne

Écrire une fonction qui recherche une chaîne dans une autre. Renvoyer le plus petit indice auquel la sous-chaîne commence, ou -1 si la sous-chaîne n'est pas présente. Exemples :

`recherche_sous_chaine("bonjour", "jour")`	$\Longrightarrow$	3
`recherche_sous_chaine("coucou", "ou")`	$\Longrightarrow$	1
`recherche_sous_chaine("bonjour", "bnjr")`	$\Longrightarrow$	-1

4 Recherche d'une chaîne dans un fichier

Le but de cet exercice est d'écrire une partie de la commande Unix `grep`. Le principe de base de `grep` est le suivant : la commande

`grep chaine fichier1 fichier2 ...`

affiche celles des lignes des fichiers spécifiés qui contiennent la chaîne `chaine`. Voici un exemple d'utilisation de `grep` :

```
$ grep machine *.html
index.html:TP 1 : familiarisation avec les machines. <p>
index.html:<code>Makefile</code> typique se trouve sur les machines de l'UFR en
pratique.html:<h2>Accès au machines de l'UFR depuis l'extérieur</h2>
pratique.html:Vous pouvez vous loger sur les machines de l'UFR en utilisant <a
```

Écrire la commande `grep`, en utilisant la fonction de l'exercice 3. On pourra supposer que les lignes des fichiers ne font pas plus de 1000 caractères.

5 Recherche d'une chaîne dans un fichier (suite)

La commande `grep` a un certain nombre de fonctionnalités supplémentaires. Vous pouvez en implémenter quelques unes.

(a) Si on ne passe aucun fichier sur son entrée standard, `grep` lit son entrée standard. Si on passe zéro ou un fichier, `grep` n'affiche que le texte des lignes qui contiennent la chaîne, sans le précéder du nom du fichier.

(b) La commande `grep` accepte un certain nombre d'*options*. Ce sont des arguments qui commencent par le caractère `-` (signe moins) et qui sont placés avant la chaîne recherchée. Par exemple :

```
$ grep -n machine *.html
index.html:13:TP 1 : familiarisation avec les machines. <p>
index.html:29:<code>Makefile</code> typique se trouve sur les machines de l'UFR en
pratique.html:46:<h2>Accès au machines de l'UFR depuis l'extérieur</h2>
pratique.html:48:Vous pouvez vous loger sur les machines de l'UFR en utilisant <a
```

La commande `grep` est appelée avec l'option `-n`, qui signifie qu'il faut afficher le numéro de la ligne entre le nom du fichier et le texte de la ligne. Modifier votre `grep` pour qu'il accepte l'option `-n`.

(c) Si on veut rechercher une chaîne qui commence par un signe moins, précéder immédiatement la chaîne par l'option `-e` pour qu'elle ne soit pas prise pour une option. Par exemple :

```
$ grep -e -1 *.html
graphics-doc.html:<code>array[3]</code>), ..., (<code>array[2*npoints-1]</code>,
references.html:Second Edition, Prentice Hall, 1988, ISBN 0-13-110362-8,
references.html:0-13-110370-9. <br> Cet ouvrage a été traduit en français :
references.html:standard américain ANS X3.159-1989. Il existe une version plus
```

Modifier votre `grep` pour qu'il accepte l'option `-e`.

(d) L'option `-i` signifie qu'il faut ignorer la différence entre majuscules et minuscules. Écrire une version modifiée de la fonction `recherche\_sous\_chaine` qui ignore la différence entre majuscules et minuscules, et l'utiliser pour ajouter l'option `-i` à `grep`.

Utiliser la fonction `toupper` définie dans l'en-tête `ctype.h`.

```
int toupper(int c);
```

Si `c` est une lettre minuscule, elle est convertie en majuscule. Sinon (lettre majuscule ou autre) `c` est inchangé. La fonction `toupper` accepte l'argument `EOF` ; il faut convertir un caractère en `unsigned char`, comme pour `isspace` et consorts (voir TP 5).

Remarque En fait la commande `grep` ne recherche pas exactement une chaîne mais un *motif*, par exemple le motif `a.*a` recherche toutes les lignes qui contiennent au moins deux fois la lettre `a`. Coder les motifs demandent des connaissances théoriques (automates finis) qui dépassent le programme de ce semestre.