

TNML302A&B

Patrick Paroubek

LIMSI-CNRS
Dépt. CHM - Groupe ILES
Rue John von Neumann, Campus Universitaire d'Orsay
Bât 508, 91405 Orsay cedex
pap@limsi.fr

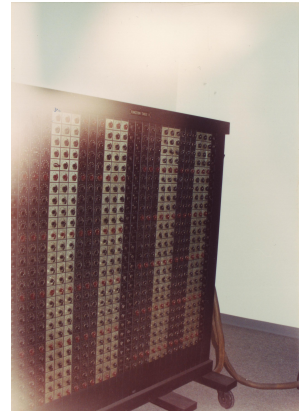
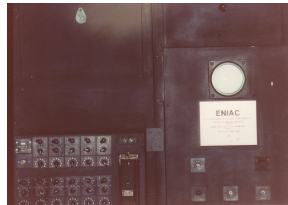
mercredi 21 septembre 2016 / TNML302A&B - Cours 1

L'informatique est une science "jeune".

- construction à la Moore School of Electrical Engineering de l'Université de Pennsylvanie à partir de mai 1944 du premier ordianteur, l'ENIAC
([http ://www.seas.upenn.edu/about-seas/eniac/](http://www.seas.upenn.edu/about-seas/eniac/))
- FTP (File Transfer Protocole), première spécification 1971
- Gopher (annuaire de téléchargement), 1991
- 1993 premier navigateur WEB (Mosaic NCSA) grand public

Naissance de l'informatique

ENIAC - 13 février 1946 - *Moore School of Elec. Engineering U Penn*



- Tout système d'exploitation dispose d'un "interprète de commande", un programme qui lit du texte au clavier, l'affiche dans un fenêtre, interprète le texte comme une commande suivie de ses arguments, et fait exécuter la commande par le système d'exploitation (**Operating System** ou **OS**), avant d'afficher les résultats de la commande.
- L'information est stockée dans des **fichiers** (*files*) qui contiennent des suites d'octets.
- Les fichiers sont groupés dans des boîtes appelées **répertoires** (ou *directories*). Un répertoire peut aussi contenir d'autres répertoires.

- Les noms de fichier peuvent être suffixés arbitrairement (parfois par plusieurs suffixes) pour indiquer leur type.
- Le caractère séparateur de suffixe est le “.” (point) aussi bien pour MSDOS qu’UNIX.
- Il est conseillé d’éviter les caractères espace (ou tabulation dans les noms de fichier) pour des raisons de portabilité.

- Commandes UNIX :
 - **cd** pour se déplacer dans l'arborescence des répertoires,
 - **ls** pour afficher le contenu d'un répertoire ,
 - **pwd** pour afficher l'endroit où l'on se trouve (par rapport à la racine de l'arborescence),
 - **cat** pour afficher la sequence d'octets contenue dans un fichier,
- Commandes Windows (anciennement MSDOS)
correspondantes : **cd, dir, pwd, type**

Pas de par cœur ! Tout système de commande a toujours 2 commandes d'information et de documentation : **help** (aide)et **man** (manuel)

- **help** donne la liste des commandes que le système reconnaît
- **man** avec comme argument le nom d'une commande, décrit en détail la sémantique et l'utilisation de cette commande.

Le répertoire courant peut aussi se noter “.”, par ex. `> cd .`, le répertoire ancêtre du repertoire courant (au dessus, en remontant vers la racine de l’arborescence) “..”, par ex. `> cd ..`.
Un chemin est dit

- "absolu", s’il est complet depuis la racine, ex.
/home/tnml3/Bureau
- "relatif", s’il utilise par défaut le repertoire courant ou son ancêtre, ex. **../Bureau**

Les chaînes de traitement

Les tuyaux

- Avec l'interprète de commandes UNIX, qui est particulièrement adapté au traitement de corpus, car toutes les commandes attendent en entrée un ou plusieurs flux de caractères et produisent en sortie un flux de caractères.
- la commande **cat** une source de caractères (voir cours 1),
- bien-sûr les commandes peuvent s'enchaîner en les reliant par des «tuyaux» *pipes*, dont le symbol est le caractère | .
- la commande > agit comme un puit de caractères en rangeant les caractères reçus sur son flux d'entrée dans un fichier dont le nom est donné en argument, dans l'ordre de leur arrivée.

Pour créer une source de caractères (flux)

- **cat**, vide le contenu du fichier argument sur le flux standard de sortie

Création d'un fichier vide :

```
>echo "" > un_fichier_vide.txt
```

Suppression d'un fichier (rm = remove)

```
>rm un_fichier_vide.txt
```

Attention :

```
> echo "aaa" > f1.txt
> cat f1.txt
aaa
> echo "bbb" > f1.txt
> cat f1.txt
bbb
```

mais

```
> echo "aaa" > f1.txt
> cat f1.txt
aaa
> echo "bbb" >> f1.txt
> cat f1.txt
aaa
bbb
```

Pour créer un repertoire

```
> mkdir mon_repertoire  
> ls -l  
drwxrwxr-x 2 pap      pap      40 Oct  7 08:36 mon_repertoire
```

Pour supprimer un repertoire

```
> rmdir mon_repertoire
```

Attention, on ne peut supprimer un repertoire que s'il est vide.

```
>mkdir mon_second_repertoire  
> cd mon_second_repertoire  
> echo "hello" > mon_fichier  
> cd ..  
> rmdir mon_second_repertoire  
rmdir: failed to remove 'mon_second_repertoire': Directory not empty
```

Il faut d'abord supprimer les fichiers qu'il contient avant de pouvoir executer `rmdir`.

```
>rm mon_second_repertoire/*  
>rmdir mon_second_repertoire
```

ATTENTION la suppression de fichiers avec un patron `**` peut vous emmener loin (pas de sauvegarde des fichiers supprimés, ils sont irrémédiablement perdus).

Les commandes sont des ordres, c'est à dire un verbe à l'impératif (le premier mot), suivit de ses arguments, s'il y en a, par exemple un COD (complément d'objet direct), un COI (complément d'objet indirect). Ex :

- `echo`
- `echo ``foobar```
- `echo -e ``foobar\nseconde ligne```

Les commandes

Syntaxe

Attention, toutes les commandes sont sensibles à la typographie au caractère près ! En particulier, dans la plupart des cas, les minuscules ne sont pas équivalentes, aux majuscules ; les espaces ont une signification et les marqueurs de portées lexicales, par ex. (),[], " ...", doivent être appariés.

Exercice.

Télécharger un fichier du web, ouvrir un interprète de commande shell (par ex. Konsole, bash, shell, xterm, ...), aller à l'endroit où le fichier récupéré se trouve, et afficher son contenu avec la commande **cat**. Puis produire la liste des mots avec leur nombre d'occurrences, triée par ordre décroissant

- entrée :

`http://abu.cnam.fr/cgi-bin/donner_unformatted?espece1`

- sortie :

- 9616 de
- 6787 les
- 5439 la
- 4756 des
- 4379 et
- 4101 que
- etc.

Ouvrir un navigateur WEB, sauvegarder la page au format html.

Problème : Trouver comment positionner le terminal pour que le répertoire courant soit le repertoire où a été sauvegardé le fichier.

Solution : utiliser les chemins apparaissant dans les barres de menu décorant les fenêtres, au besoin utiliser un outil de recherche du fichier, à la condition d'avoir donné un nom caractéristique au fichier dans lequel vous avez sauvegardé votre page web.

Pour afficher un mot par ligne de votre fichier **monFichier.html**.

```
cat monFichier.html | tr '\040' '\012'
```

Pour trier les mots par ordre alphabétique

```
cat monFichier.html | tr '\040' '\012' | sort
```

Pour ceux qui n'ont pas de Linux installé sur leur machine et qui ne veulent pas installer une machine virtuelle sur leur PC, ils peuvent installer **cygwin**, un émulateur Linux qui permet de disposer d'un terminal de commandes Linux.

<https://www.cygwin.com/>

Install it by running `setup-x86.exe` (32-bit installation) or `setup-x86_64.exe` (64-bit installation).