

## TD n°4 – expressions régulières

Soit le fichier `texte` contenant ceci (les numéros de lignes mis dans la marge ne sont pas dans le fichier) :

1	Titre
2	Android est un systeme d'exploitation destine aux tablettes
3	tactiles. Il est ne le 05/11/2007 et plusieurs versions ont vu le
4	jour depuis. La derniere est la 5.1 et une des plus utilisees est
5	la 4.4. Voici les chiffres :
6	
7	Version 5.1 : 5.1% des machines
8	Version 5.0 : 16% des machines
9	Version 4.4 : 39% des machines

NB: la ligne 6 est entièrement vide. D'autre part, il n'y a qu'un seul espace seulement là où on en voit ; il n'y en a ni en début de ligne, ni en fin de ligne. Il n'y a pas d'accents pour ne pas compliquer les recherches.

NB: ces statistiques sont biaisées, car basées sur les connexions aux Google Play Store sur les 7 jours précédant le relevé. Cela ne représente donc absolument pas l'état réel du marché, mais seulement les machines dont les propriétaires ont installé ou mis à jour un nouveau logiciel, ce qui est certainement plus fréquent avec une tablette neuve qu'avec une plus ancienne.

Pour faire les tests, ce qui peut être pratique, c'est de taper `alias egrep='egrep --color=auto'`  
Également on peut rajouter l'option `-n` à `egrep` pour afficher les numéros des lignes sélectionnées.

### 1) Partie 1

Pour chaque question qui suit, on demande les numéros des lignes affichées par les commandes avec une justification : pourquoi ce sont ces lignes là qui sont sélectionnées (que signifie le motif) :

- a) `egrep '[A-Z]a' texte`
- b) `egrep '[a-z]e$' texte`
- c) `egrep '[oiseau][vole]' texte`
- d) `egrep '^ [A-Z][a-z]' texte`
- e) `egrep '^[A-Z]o' texte`
- f) `egrep '[ ]+[a-z]{6,8} ?\.'` texte
- g) `egrep -v '^[ ]' texte`
- h) `egrep '^[^ ]+$' texte`
- i) `egrep '^[emiaou]{2}$' texte`
- j) `egrep '[0-9]+[^0-9a-z]' texte`
- k) `egrep '([0-9])\.\1' texte`

### 2) Partie 2

Maintenant à vous de proposer des expressions régulières pour rechercher selon les critères suivants :

- l) les lignes qui contiennent le mot *version* en majuscules ou minuscules
- m) les lignes qui finissent par le mot *le*.
- n) les lignes qui contiennent les mots *le* ou *la*. (2 solutions).
- o) les lignes qui contiennent les mots *un* ou *une*. (2 solutions).
- p) Les lignes qui contiennent deux fois le mot *le*.
- q) les lignes qui contiennent un mot en minuscule suivi d'un point.

- r) les lignes qui contiennent un nombre entier suivi du signe %. Comment empêcher de reconnaître un nombre à virgule ?
- s) les lignes qui contiennent un numéro de version correctement écrit (à définir).
- t) les lignes qui contiennent une date écrite sous la forme de chiffres jour/mois/année, en utilisant au mieux les jokers.
- u) Les lignes qui contiennent un nombre entier compris entre 1 et 19.
- v) Les lignes qui contiennent la même minuscule répétée deux fois à la suite.
- w) Les lignes qui contiennent deux fois le même nombre à virgule.