## Exercice 1

<instance id="d005.s011.t001" lemma="drawing" pos="n">drawings</instance>  
 <instance id="d005.s011.t002" lemma="improve" pos="v">improved</instance>  
 <instance id="d005.s011.t003" lemma="careless" pos="a">careless</instance>  
 <instance id="d005.s011.t004" lemma="habit" pos="n">habits</instance>  
 <instance id="d005.s011.t005" lemma="continue" pos="v">continued</instance>

### indiquez les annotations fournies .

### <instance id="d005.s011.t001" lemma="drawing" pos="n">drawings</instance>

* Position : document 5, 11ème phrase, 1er terme
* lemme : drawing
* terme : drawings

### <instance id="d005.s011.t002" lemma="improve" pos="v">improved</instance>

* Position : document 5, 11ème phrase, 2eme terme
* lemme : improve
* terme : improved

### <instance id="d005.s011.t003" lemma="careless" pos="a">careless</instance>

* Position : document 5, 11ème phrase, 3eme terme
* lemme : careless
* terme : careless

### <instance id="d005.s011.t004" lemma="habit" pos="n">habits</instance>

* Position : document 5, 11ème phrase, 4eme terme
* lemme : habit
* terme : habits

### <instance id="d005.s011.t005" lemma="continue" pos="v">continued</instance>

* Position : document 5, 11ème phrase, 5eme terme
* lemme : continue
* terme : continued

### Quels sont les quatre parties du discours des mots annotés ?

* N = nom
* A = adjectif
* R = adverbe
* V = verbe

### Pourquoi certains mots ne sont pas annotés ?

Parce qu’ils ne présentent pas d’ambiguïté.

## Exercice 2

### Pourquoi y-a-t-il plusieurs synsets possible dans certain cas ?

Il y a plusieurs synsets car, le dictionnaire peut renvoyer plusieurs sens à chaque mot dans cette phrase.

## Exercice 3

### Pour chacun des exemples, donnez la précision, le rappel et le F-score.

* Fichier 1

Total: A = 1.000 (2269 / 2269), P = 0.74394 (1688 / 2269), R = 0.74394 (1688 / 2269), F1 = 0.74394

* Fichier 2

Total: A = 1.000 (2269 / 2269), P = 0.73601 (1670 / 2269), R = 0.73601 (1670 / 2269), F1 = 0.73601

* Fichier 3

Total: A = 1.000 (2269 / 2269), P = 0.75187 (1706 / 2269), R = 0.75187 (1706 / 2269), F1 = 0.75187

* Fichier 4

Total: A = 1.000 (2269 / 2269), P = 0.62054 (1408 / 2269), R = 0.62054 (1408 / 2269), F1 = 0.62054

* Fichier 5

Total: A = 0.859 (1948 / 2269), P = 0.76027 (1481 / 1948), R = 0.65271 (1481 / 2269), F1 = 0.70240

A = attemped : Nombre de mots non annotés / nombre de mots.  
P = précision : Nombre de mots bien annotés / nombre de mots annotés   
R = rappel : Nombre de mots bien annotés / nombre de mots.  
F1 = F. Mesure

## Exercice 4

-w affiche les mots mal annotés  
-m affiche la référence des mots manquants  
-d affiche l'analyse de tous les doccuments du dossier  
-x exclude one-cluster words  
-a affiche toutes les réponses de chaque terme

## Exercice 5

## Exercce 6

### Quelle sont les deux premières définitions ?

* [S:](http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=cat&i=0&h=0000000000#c) (n) **cat**, [true cat](http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=true+cat) (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
* [S:](http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=cat&i=1&h=0000000000#c) (n) [guy](http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=guy), **cat**, [hombre](http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=hombre), [bozo](http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=bozo), [sod](http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=sod) (an informal term for a youth or man) *"a nice guy"; "the guy's only doing it for some doll"; "the poor sod couldn't even buy a drink"*

On affecte un entier à chacun des mots

Feline = 1  
mammal = 2  
usually = 3  
having = 4  
thick = 5  
soft = 6  
fur = 7  
and = 8  
no = 9  
ability = 10  
to = 11  
roar =12  
domestic = 13  
cats =14  
wildcats = 15  
an =16  
informal =17  
term = 18  
for =19  
a =20  
youth 21   
or =22  
man =23

Ensuite, on regroupe les ensembles d’entiers qui codent les mots, par définitions

D1->{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15}  
D2 ->{16, 17, 18, 19, 20, 21, 22, 23}

Ces ensembles sont tries par ordre croissant

## Exercice 7

<ids>cat%1:05:00:: </ids>  
<def>8 22 112 428 528 705 1216 1228 1538 2450 2518 7518 9842 31822 45730 </def>

<ids>cat%1:18:00:: </ids>  
<def>0 0 72 134 409 482 10357 13768 16164 </def>

On peut voir que la seconde définition présente deux fois le même mot (0). Ce mot doit correspondre à « a » et « an » que l’on a considéré comme différents précédemment.

## Exercice 8 à 10

Voir fichier java.

## Exercice 11

Function getSimilarity1

Similarité entre pine et cone sur le dictionnaire Dict-Lesk.xml : 0.8

Similarité entre pine et cone sur le dictionnaire Dict-Lesk-etendu.xml : 11.533334

Similarité entre pine et cone sur le dictionnaire dict\_all\_stopwords\_stemming\_semcor\_dso\_wordnetglosstag\_150.xml : 4.4666667

## Exercice 12

media,radio,7.42

tiger,tiger,10.00

stock,jaguar,0.92

football,soccer,9.03

brother,monk,6.27

Le fichier contient une liste de couples de mots, associée à un nombre compris entre 0 et 10. Ce nombre représente la similarité entre ces deux mots : on peut voir que, entre deux fois le même mot, la similarité est de 10 (tiger, tiger), alors que entre deux mots qui n’ont rien à voir (stock, jaguar), la similarité est très basse.

## Exercice 13

La corrélation entre les séries {1,4,8,12,2,5,4} et {4,2,1,6,6,10,4} est de 0.004243272981903036.

## Exercice 14

Utilisation de OpenCSV  
Corrélation entre humain et lesk :0.20313210443967666

Corrélation entre humain et lesk étendu :0.21253348768870164

Corrélation entre humain et lesk corpus :0.21388841987844234

## Exercice 16

36 combinaisons pour la chaine "go" dans le 1e dictionnaire

96 combinaisons pour la chaine "mouse pilot computer" dans le 1e dictionnaire

13440 combinaisons pour la chaine "dog eat bone every day" dans le 1e dictionnaire

470400 combinaisons pour la chaine "doctor be hospital last day night" dans le 1e dictionnaire

7832809874592301056 combinaisons pour la chaine "pictures paint be flat round figure be very often foot do look be stand ground all point downward be hanging air" dans le 1e dictionnaire

36 combinaisons pour la chaine "go" dans le 2e dictionnaire

96 combinaisons pour la chaine "mouse pilot computer" dans le 2e dictionnaire

13440 combinaisons pour la chaine "dog eat bone every day" dans le 2e dictionnaire

470400 combinaisons pour la chaine "doctor be hospital last day night" dans le 2e dictionnaire

7832809874592301056 combinaisons pour la chaine "pictures paint be flat round figure be very often foot do look be stand ground all point downward be hanging air" dans le 2e dictionnaire

36 combinaisons pour la chaine "go" dans le 3e dictionnaire

96 combinaisons pour la chaine "mouse pilot computer" dans le 3e dictionnaire

13440 combinaisons pour la chaine "dog eat bone every day" dans le 3e dictionnaire

470400 combinaisons pour la chaine "doctor be hospital last day night" dans le 3e dictionnaire

7832809874592301056 combinaisons pour la chaine "pictures paint be flat round figure be very often foot do look be stand ground all point downward be hanging air" dans le 3e dictionnaire

## Exercice 17

La première et la dernière phrase ne semblent pas désambiguisable. En effet, la première ne présente pas de contexte, et la dernière présente beaucoup trop de combinaisons pour pouvoir toutes les tester.

## Exercice 19

« mouse pilot computer » :