TP1
Map Reduce simplifié

Objectif : Compter des éléments dans un dictionnaire à partir d'un texte à analyser.

1. Composer le dictionnaire de manière à ce que chaque mot différent du texte constitue une entrée.
2. Chaque entrée permet de dénombrer le nombre d'occurrences de chaque mot.
3. Trier le dictionnaire avec les entrées les plus nombreuses en haut.
4. Proposer une visualisation des données avec un histogramme en barres horizontales.

Texte de base :

Le petit poucet c'est perdu dans la campagne ! Il a peur … A vous de l'aider. Pas facile ? Le petit poucet compte sur vous !

Consignes :

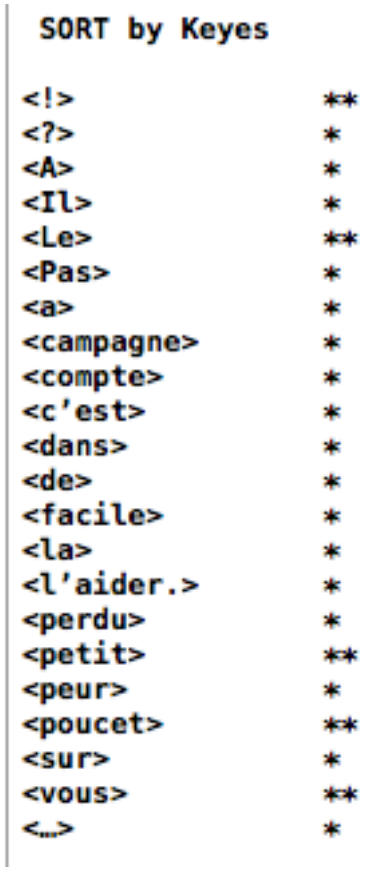
Chaque occurrence sera représentée par une étoile :

Non trié :

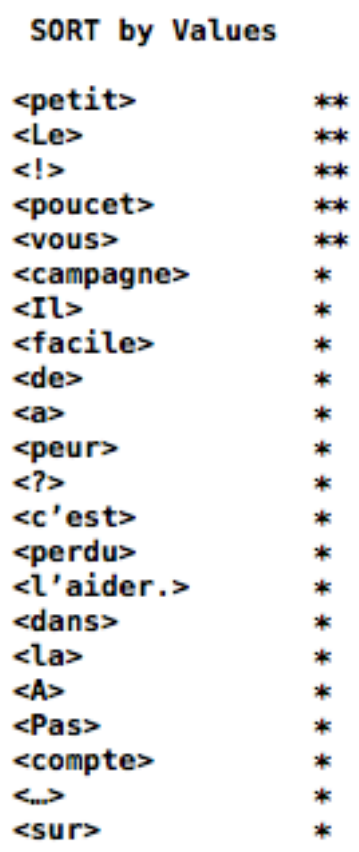
```
 UNSORT

<campagne>        *
<l'aider.>        *
<!>               **
<?>               *
<>                *
<de>              *
<facile>          *
<Il>              *
<petit>           **
<vous>            **
<a>               *
<peur>            *
<poucet>          **
<c'est>           *
<perdu>           *
<Le>              **
<dans>            *
<la>              *
<A>               *
<Pas>             *
<compte>          *
<…>               *
<sur>             *
```

Après tri sur les clés :



Après tri sur les valeurs :



Consignes autour du projet : Le projet sera réalisé dans une console pour faire simple. La version de base doit pouvoir supporter au moins le jeu d'essai sur une œuvre d'Harry Potter.

Vous pouvez vérifier votre projet avec l'intégrale de « Harry Potter à l'école des sorciers ! »

Evidemment, pour vérifier la régularité du traitement, il va faudra lire l'œuvre intégrale :)

Le mot « de » apparaît 3226 fois `<1><de><3226>` et le mot « Harry » 652 fois `<17><Harry><652>` ! Le traitement génère 15021 entrées dans le dictionnaire.

`<15021><courageux...><1> *`

Attention, le mot Harry apparaît également dans d'autres formats, « Harry. » « Harry, » « ? Harry » … Il s'agit donc de valeurs brutes à retravailler pour obtenir une vraie tendance, sur les répétitions. Mais cela n'était pas l'objectif de cette version.

```
<166><beaucoup><57>      ************************************************************
<167><quelques><57>      ************************************************************
<168><sais><56>          ***********************************************************
<169><toute><56>         ***********************************************************
<170><«><56>             ***********************************************************
<171><qu'elle><55>       *********************************************************
<172><baguette><55>      *********************************************************
<173><trop><55>          *********************************************************
<174><pouvait><55>       *********************************************************
<175><pu><54>            ********************************************************
<176><dit-il><54>        ********************************************************
<177><mon><54>           ********************************************************
<178><trois><53>         *******************************************************
<179><ces><53>           *******************************************************
<180><prendre><53>       *******************************************************
<181><vraiment><51>      *****************************************************
<182><autre><51>         *****************************************************
<183><leurs><51>         *****************************************************
<184><m'a><51>           *****************************************************
<185><sous><51>          *****************************************************
<186><aller><50>         ****************************************************
<187><vu><50>            ****************************************************
<188><Gryffondor><49>    ***************************************************
<189><autour><49>        ***************************************************
<190><long><49>          ***************************************************
<191><veux><49>          ***************************************************
<192><j'ai><49>          ***************************************************
<193><avez><49>          ***************************************************
<194><A><48>             **************************************************
<195><cape><47>          *************************************************
<196><semblait><47>      *************************************************
<197><regard><47>        *************************************************
<198><as><47>            *************************************************
<199><quelqu'un><46>     ************************************************
```