

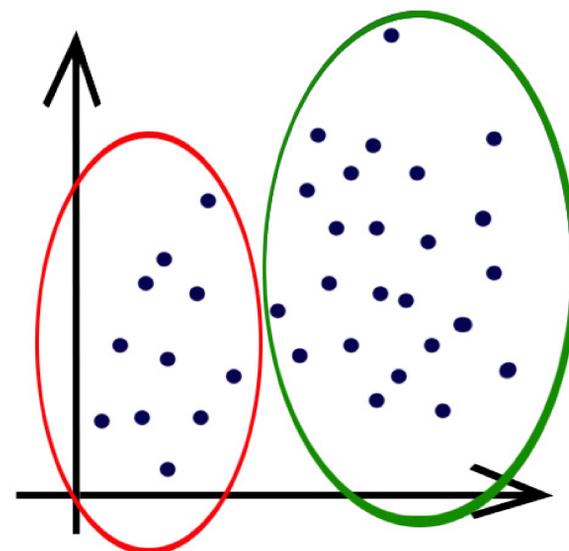
## Clustering de Musique

Boris Iolis, Svetlana Lukitcheva, Nikita Veshchikov  
Département d'Informatique

### Clustering:

Répartir un ensemble de données en groupes, de manière à ce que chaque groupe contienne des données similaires.

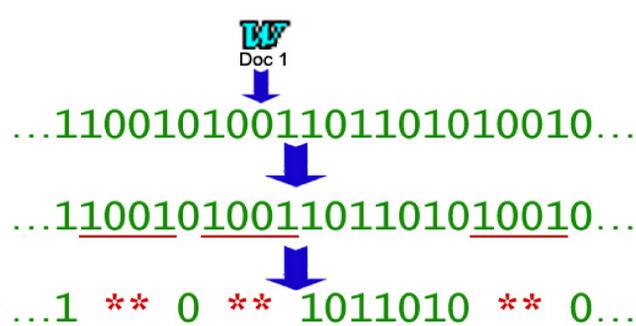
Dans notre cas, les données à grouper seront des morceaux de musique. Ces morceaux devront donc être groupés par genre musical.



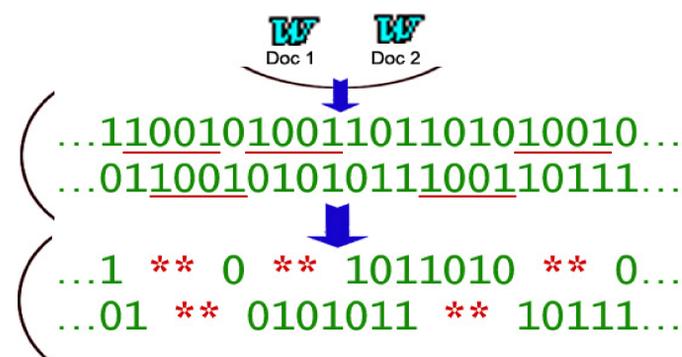
Comment savoir quels morceaux sont similaires?

Grâce à un algorithme de compression.

Les algorithmes de compression recherchent des patterns (motifs) qui se répètent dans un document.



\*\* = 1001



\*\* = 1001

En compressant ensemble deux fichiers musicaux, on peut donc en déduire leur similitude :

$$\text{Similitude}(a, b) = \frac{\text{Comp}(a) + \text{Comp}(b)}{2 \times \text{Comp}(ab)}$$

où  $\text{Comp}(a)$  est la taille du fichier 'a' une fois compressé, et  $\text{Comp}(ab)$  est celle des fichiers 'a' et 'b' compressés ensemble