



### **Why**

We want to talk about categorizing documents.

### **Definition**

Let  $X$  be a set of documents. To be concrete, for  $n \in \mathbf{N}$ ,  $X$  may be all length  $n$  character sequences in some alphabet of digital symbols.



