**Term Frequency-Inverse Document Frequency**

- How important a word is to a document $D$ in a collection $C$ ($|C| = N$)?

- Term frequency
  
  $tf_{t,D} = \text{the number of times a term } t \text{ occurs in } D$

- Document frequency
  
  $df_{t,D} = \text{the number of documents in } C \text{ in which a term } t \text{ occurs, i.e.,}$
  
  $|\{D \in C : t \in D\}|$

- Inverse document frequency, term importance
  
  $idf_{t,D} = \log \frac{N}{df_{t,D}}$

  $tfidf_{t,D,C} = tf_{t,D} \cdot idf_{t,C}$