# 🤓 Information Retrieval (RI) – TPs

---

## ◇ TF (Term Frequency)

**Definition:**
TF measures how frequently a term appears in a document.

**Formula:**

**TF(t, d) = (Number of times term t appears in document d) / (Total number of terms in document d)**

**Intuition:**
Common words in a document get higher TF values.

**Example:**
Document: “the cat sat on the mat”
→ TF("cat") = 1 / 6 = **0.1667**

---

## ◇ IDF (Inverse Document Frequency)

**Definition:**
IDF measures how unique or rare a term is across all documents in a corpus.

**Formula:**

**$IDF(t) = \log( N / df_t )$**

Where:

- **N** = total number of documents
- **$df_t$** = number of documents containing term *t*

**Intuition:**

- Words appearing in many documents (like “the”, “and”, “is”) get **low IDF** (less informative).
- Rare words get **high IDF** (more informative).

---

## ◇ TF-IDF (Term Frequency – Inverse Document Frequency)

**Definition:**
TF-IDF combines both TF and IDF to measure how important a term is to a document in a collection.

**Formula:**

**TF-IDF(t, d) = TF(t, d) × IDF(t)**

**Intuition:**
High when a term is frequent in a document but rare in the corpus.

Helps identify keywords that best represent each document.