$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \times \text{idf}(t, D)$$

	Doc 1	Doc 2	...	Doc *n*
Term(s) 1	12	2	...	1
Term(s) 2	0	1	...	0
...	...	...	...	
Term(s) *n*	0	6	...	3

- A **TF-IDF** measure (**term-frequency inverse [TF] document frequency [IDF]**) is a commonly used counting technique intended to reflect how relevant a term is in a given document.

- In other words, it is a **document specific measure** that tells you how important an **n-gram** is in a document **relative** to a **corpus**.

$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \times \text{idf}(t, D)$$

	Doc 1	Doc 2	...	Doc *n*
Term(s) 1	12	2	...	1
Term(s) 2	0	1	...	0
...	...	...	...	
Term(s) *n*	0	6	...	3

- The intuition behind it is that if an **n-gram** occurs multiple times in a document, we should boost the overall relevance of the **n-gram**.

- In other words, it should be more **meaningful** than other words that appear fewer times. This is what the **TF** measures.

- In other words, if should be more **meaningful** than other words that appear fewer times. This is what the **TF** measures.

of the **n-gram**.

- The intuition behind it is that if an **n-gram** occurs multiple times in a document, we should boost the overall relevance