

Similarity Metrics for Medical Text



Cosine Similarity

$$\cos(\theta) = \frac{A \cdot B}{(\|A\| \|B\|)}$$

Range: [-1, 1]

Best for: Dense embeddings



Euclidean Distance

$$d = \sqrt{\sum (a_i - b_i)^2}$$

Range: [0, ∞]

Best for: Spatial similarity



Jaccard Index

$$J = \frac{|A \cap B|}{|A \cup B|}$$

Range: [0, 1]

Best for: Set overlap



Semantic Similarity

Based on medical ontology

Range: [0, 1]

Best for: Medical concepts



Medical Text Example

Text 1: "Patient has myocardial infarction"

Text 2: "Heart attack diagnosed"

Cosine: **0.89** ✓

Jaccard: **0.12** ✗

Semantic: **0.95** ✓