PTC | POLYTECHNIC COLLEGE SURINAME

UNIVERSITY OF APPLIED SCIENCES

QUALITY SYSTEM CERTIFICATION DNV ISO 9001

Evaluatie en feedback sessie [Recap afgelopen 5 weken]

Tuesday, November 18, 2025

PTC | POLYTECHNIC COLLEGE SURINAME

UNIVERSITY OF APPLIED SCIENCES

QUALITY SYSTEM CERTIFICATION DNV ISO 9001

Similarity Queries and Summarization

Tuesday, November 18, 2025

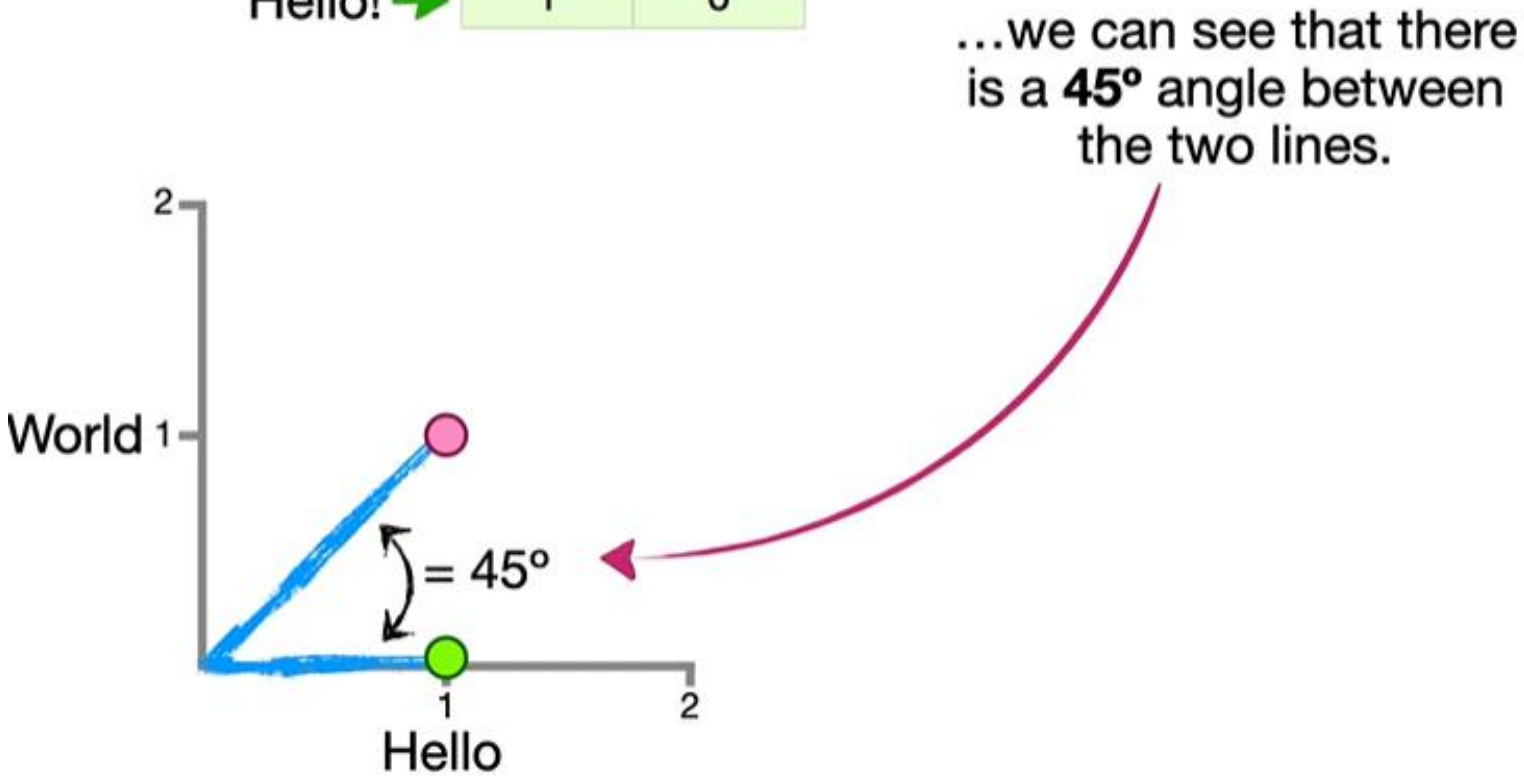
Cosine Similarity

The Cosine Similarity is a relatively easy to calculate metric that can tell us how similar or different things are

- Measures how similar two texts (or vectors) are
- Based on the angle between two vectors
- Ranges between 0 (completely different) and 1 (identical)
- Commonly used in text analysis, NLP, and search engines

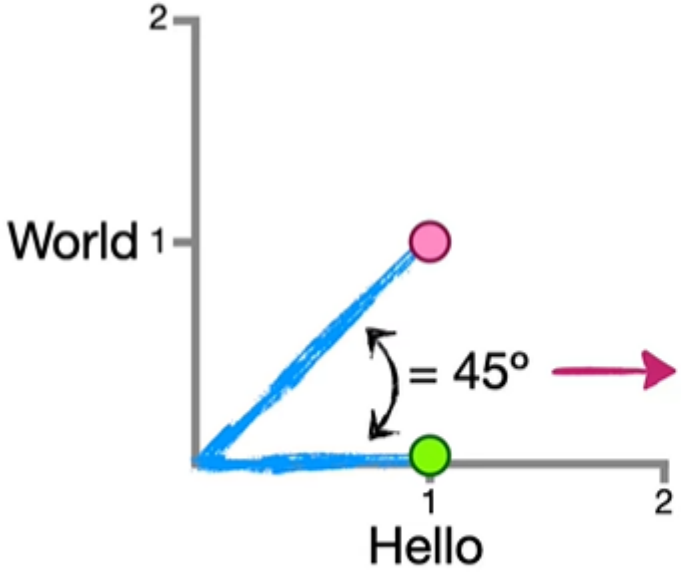
Cosine Similarity

	Hello	World
Hello, World! →	1	1
Hello! →	1	0



Cosine Similarity

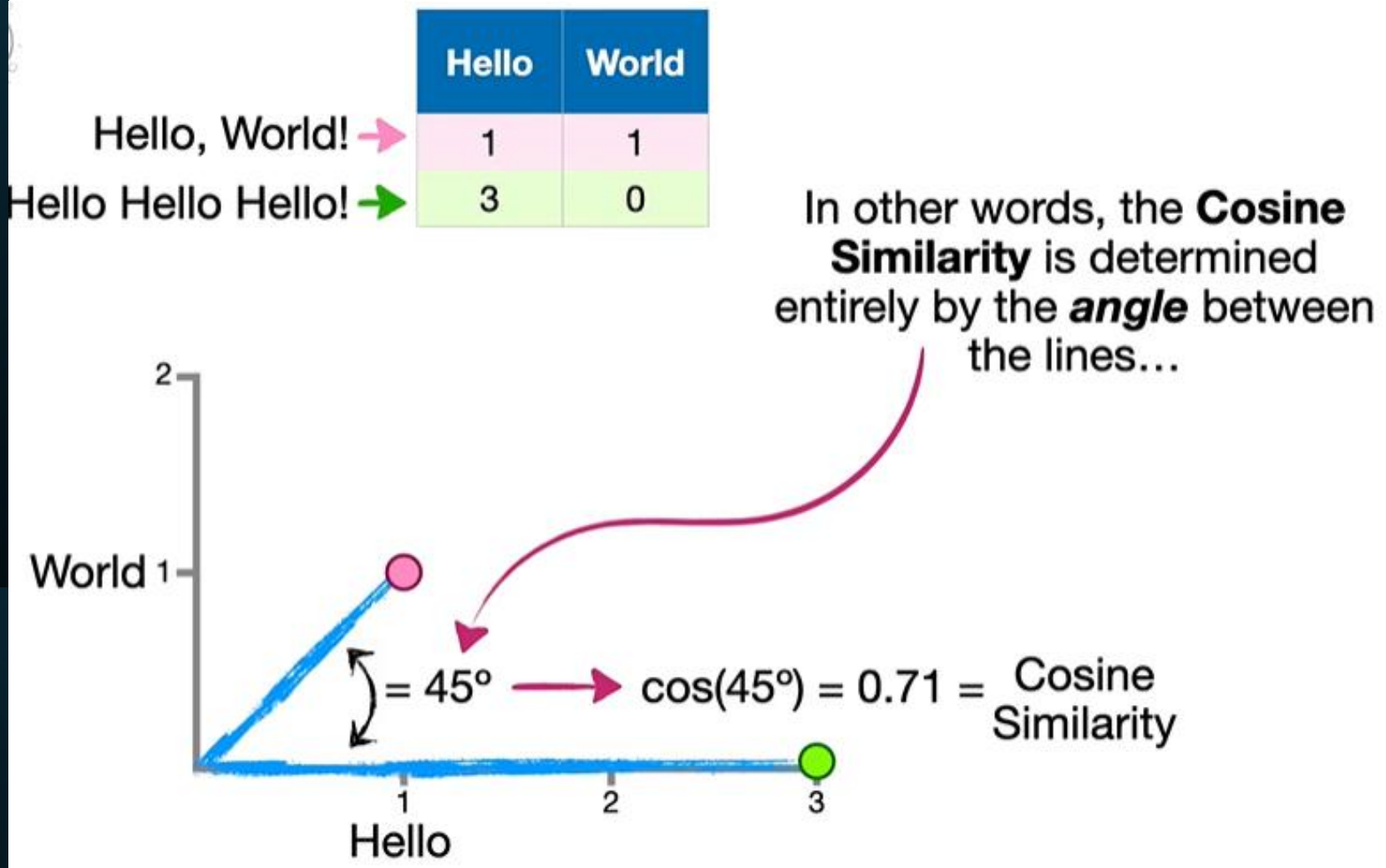
	Hello	World
Hello, World! →	1	1
Hello! →	1	0



…and thus, the **Cosine Similarity** between the phrases, "**Hello, World!**" and "**Hello!**" is **cos(45°) = 0.71**.

cos(45°) = 0.71 = Cosine Similarity

Cosine Similarity

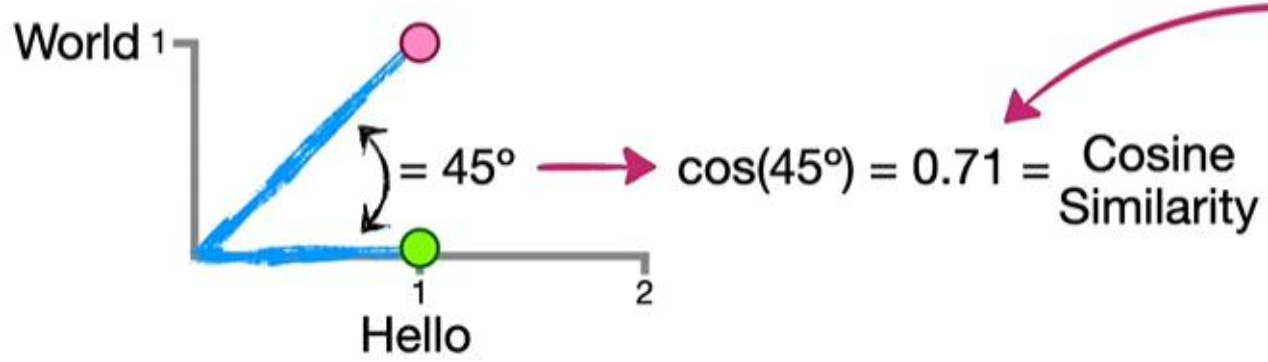


Cosine Similarity

	Hello	World
A = Hello, World! →	1	1
B = Hello! →	1	0

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} = \frac{(1 \times 1) + (1 \times 0)}{\sqrt{1^2 + 1^2}\ \sqrt{1^2 + 0^2}} = 0.71$$

…which is the same value we got when we took the **cos(45°)**.



Cosine Similarity

	I	Love	Troll	2	Gymkata
A = I love Troll 2! →	1	1	1	1	0
B = I love Gymkata! →	1	1	0	0	1

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Then we do the math and we get **0.58**.

$$= \frac{(1 \times 1) + (1 \times 1) + (1 \times 0) + (1 \times 0) + (0 \times 1)}{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 0^2}\sqrt{1^2 + 1^2 + 0^2 + 0^2 + 1^2}}$$

$$= 0.58$$

Text Summarization

- Condenses long text into shorter version
- Saves time while still getting the essence of the content
- Used by Journalists, Students, Businesses, etc.

Hugging Face Transformers

- State-of-the-art platform for NLP
- Pre-trained models of translation, sentiment, analysis, summarization, etc.
- Understands and generates language in a human-like way

Hugging Face Transformers

- https://huggingface.co/facebook/bart-large