ROUGE-N (Recall-Oriented Understudy for Gisting Evaluation):

a metric within the broader ROUGE metric collection, is a vital metric in the field of natural language processing and text evaluation. It assesses the quality of a candidate text by measuring the overlap of n-grams between the candidate text and reference texts, and ranges between 0 and 1. A score of 0 indicates no overlap between candidate and reference texts, whereas a perfect score of 1 indicates perfect overlap. ROUGE-N provides insights into the ability of a system to capture essential content and linguistic nuances, making it an important and versatile tool used in many NLP workflows. As the name implies, it is a recall-based metric — a complement to the precision-based BLEU score.

ROUGE-1 is a metric used to evaluate the **quality of generated text**, especially in tasks like **text summarization**, **machine translation**, and **text generation**. It focuses on **unigram (single word) overlap** between a **generated (candidate) text** and one or more **reference texts** written by humans.

Example:

Reference Text:

"The cat sat on the mat."

Candidate Text:

"The cat is sitting on the mat."

Step 1: Extract Unigrams (ROUGE-1)

Unigrams are individual words.

Reference unigrams:

the, cat, sat, on, the, mat
(Note: "the" appears twice)

Candidate unigrams:

the, cat, is, sitting, on, the, mat

Step 2: Count Overlapping Unigrams

Overlapping unigrams between reference and candidate:

- "the" (counts as 2, since both have it twice)
- "cat"
- "on"
- "mat"

That's **5 overlapping unigrams**.

Step 3: Calculate ROUGE-1 Recall

ROUGE-N is **recall-based**, so the formula is:

ROUGE-1 Recall=Overlapping unigrams / Total unigrams

ROUGE-1 Recall=5/6= =0.8333

Final ROUGE-1 Score: 0.83

This shows that the candidate sentence covers about 83% of the content (unigrams) in the reference, even though the wording is slightly different. This is why ROUGE-N is useful for evaluating summaries, translations, and similar tasks.