

prompt = f"""

Here are the two descriptions to compare:

Original description:

<original_description>

{sentence1}

</original_description>

Generated description:

<generated_description>

{sentence2}

</generated_description>

Please follow these steps in your analysis:

1. Carefully read both descriptions.
2. List key characteristics of each description, including:
 - Disease identification
 - Disease progression
 - Symptoms
 - Treatment implications
3. Compare the characteristics point-by-point, noting similarities and differences.
4. Consider synonymous terms (e.g., "spots" vs "lesions", "tan" vs "light brown").
5. Assess the overall disease impact and severity.
6. Determine if the fundamental disease presentation and required treatment would be the same.
7. Summarize the similarities and differences.

Based on your analysis, assign a similarity score from 0 to 1, where:

- 0.0-0.5: Different diseases or requiring different treatments
- 0.5-0.79: Same disease but may need treatment adjustments
- 0.8-1.0: Essentially the same disease and treatment approach

Before providing your final output, wrap your detailed comparison and thought process in <comparison_process> tags. Follow this step-by-step process:

- a. Extract and list key characteristics from each description
- b. Compare these characteristics side-by-side
- c. Note similarities and differences
- d. Consider synonymous terms
- e. Assess overall disease impact and severity
- f. Determine if fundamental disease presentation and treatment would be the same

This will ensure a thorough and transparent evaluation.

After your analysis, provide your response as a Python dictionary with two keys:

1. 'score': A float between 0 and 1 representing the similarity.
2. 'reason': A brief rationale for the score (maximum 15 words).

Your output should follow this structure:

<comparison_process>

[Your detailed comparison and thought process]

</comparison_process>

{{"score": [Your score as a float], "reason": "[Your brief reason for the score]"}}

Remember:

- High similarity (0.8-1.0) would be for descriptions that are essentially identical in terms of disease

and treatment approach, even if using different terminology. For example:

Original: "maize leaf blight, moderate severity symptoms: tan to brown elongated lesions"

Generated: "maize leaf blight, moderate infection symptoms: narrow, elongated sores"

- Medium similarity (0.5-0.79) would be for descriptions of the same disease but with notable differences

that may affect treatment. For example:

Original: "maize leaf blight, mild symptoms: small brown spots"

Generated: "maize leaf blight, severe symptoms: large spreading lesions"

- Low similarity (0.0-0.5) would be for descriptions that appear to be discussing different diseases

or require significantly different treatments. For example:

Original: "maize leaf blight, mild to moderate infection symptoms: brown spots, lesions with halos"

Generated: "yeast infection, moderate to severe infection Severity increased over time"

Please provide your analysis and final response now."""

system_prompt = (

"You are an expert plant pathologist specializing in maize diseases. Your task is to compare two "

"descriptions of maize leaf blight, focusing on core disease characteristics and treatment "

"implications. Your analysis should allow for synonymous terminology while maintaining accuracy in "

"disease management."

)