**Observe como el el algoritmo encuenta coincidencias basandose en la similitud del texto,**

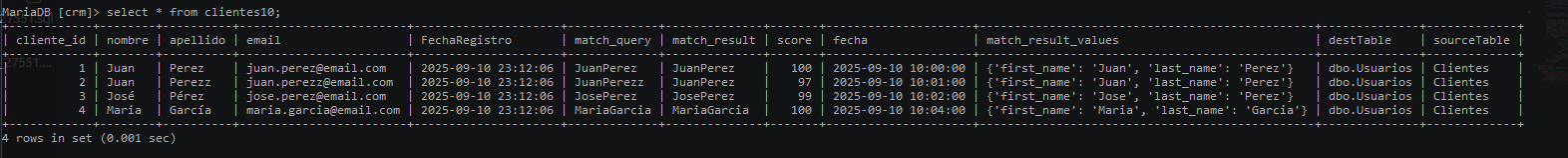
**Cree o busque ecenarios de datos para responder a las siguientes preguntas:**

1- Como se comporta el algoritmo si hay errores tipograficos en el apellido?(p.ej. "Perezz" vs "Perez")

r=Si un registro tiene "Perezz" y otro "Perez", RapidFuzz calculará un score alto (por ejemplo, 96-98%).

Si el score es 97% o más, el sistema lo considerará "coincidente" y lo insertará en la tabla Coincidentes.

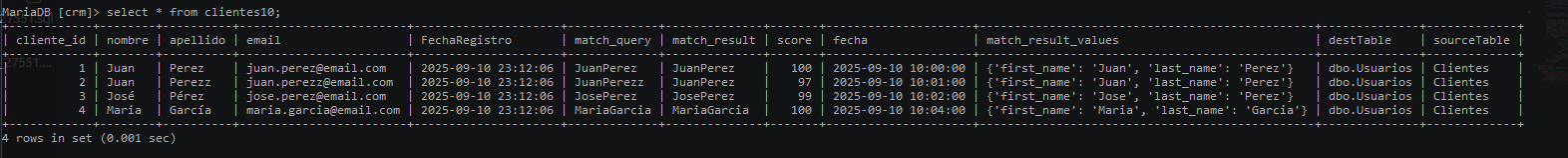
Si el score es 96% o menos, el registro será "no coincidente" y no se insertará.



2- Que pasa si faltan acentos o caracteres especiales(p. ej. "Pérez" vs "Perez")

r=Si el score es 97% o más, el sistema lo considera "coincidente" y lo inserta.

Si el score baja de 97%, no se inserta.



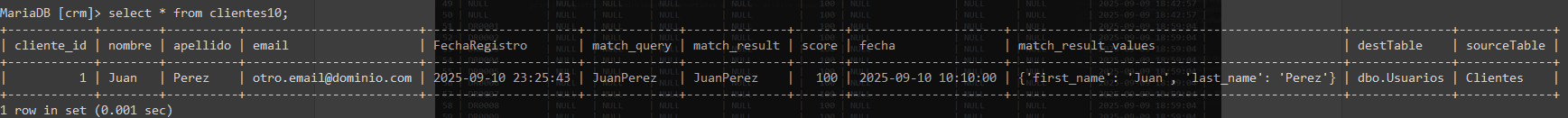
3- Si dos registros tiene el mismo "score", Como deberia el algoridmo decicir cual es la mejor coincidencia?

r=Actualmente, el sistema selecciona la primera coincidencia encontrada con el score más alto,

no hay desempate adicional: simplemente se inserta el primero que cumple el criterio.

4- Que sucede si el nombre coincide pero el email es muy diferente?

r=Si el nombre y apellido coinciden pero el email es muy diferente, el sistema puede seguir considerando el registro como coincidente si el score de similitud es suficientemente alto



5- Cuales son las limitaciones de tratar todas las columnas con la misma importacia?

r=Tratar todas las columnas con la misma importancia puede llevar a:

Falsos negativos: Si un campo poco relevante (como un segundo apellido) es diferente, puede bajar el score total y descartar coincidencias válidas.

Falsos positivos: Si campos clave (como email) coinciden pero otros no, el sistema puede aceptar coincidencias incorrectas.

Falta de flexibilidad: No todos los campos tienen el mismo peso en la identificación de una persona.

6- Puede pensar en situaciones reales (p ej., bases de datos de clientes, registros hospitalarios, comercio electronico)

en las que la coincidencia difusa podria crear problemas si todas las columnas se tratan por igual?

r=a)En bases de datos de clientes, un error en el nombre pero coincidencia en el email debería pesar más el email.

b)En registros hospitalarios, el número de seguro social o fecha de nacimiento deberían tener más peso que el nombre.

c)En comercio electrónico, el teléfono o dirección pueden variar, pero el email o ID de usuario deberían ser determinantes.

Si todas las columnas se tratan igual, se pueden unir registros de personas distintas o no detectar duplicados reales.

