**Plan de Testing**

**Síntesis:** La idea del documento es plantear los lineamientos bajo los cuales se va a testear la calidad de la comparación de textos. Definiendo casos base, dividiendo datasets de entrenamiento/prueba, framework de testing y todo lo necesario para que cada colaborador pueda ejecutar la suite de tests y evaluar si sus cambios mejoran o empeoran la performance de Comparación

Tecnologías a Utilizar:

**Lenguaje**: Python

**Metodología**: Unit Tests e Integration Tests

**Framework de Testing**: pytest

**Repositorio**: Github (público)

Lineamientos Generales

* Los Unit Tests (UT) deben ser minimos y faciles de entender/modificar. Su propósito debe ser único y probar cierta funcionalidad. (No todas!)
* Debe existir un caso de UT por cada caso de comparación
* Las variables/nombres de funciones deben ser en inglés preferentemente
* Se deben seguir dentro de lo posible los lineamientos de estilo definidos en PEP8

Casos de Test:

1. Dos textos idénticos deben dar similitud >= 99%
2. Dos textos que no comparten ninguna palabra deben dar similitud <= 10%
3. Una cita sin la referencia correspondiente debe penalizar el score final
4. El algoritmo debería comprender las distintas formas válidas de citar una referencia externa
5. Dos oraciones idénticas en textos distintos deberían aumentar la similitud
6. Casos conocidos de plagio (reales y artificales) deben dar un porcentaje mayor a 60%. Idealmente mayor a 80%.
7. Una copia parcial de una oración textual debería penalizar la oración, aunque menos que si copia la oración total.