**Criterios de éxito**

|  |  |
| --- | --- |
| Título del proyecto de tesis: | Caracterización de recursos computacionales para la fase de preprocesamiento de minería de textos |
| Criterios: | La problemática a resolver mediante este proyecto, radica en aprovechar las funcionalidades que proveen tanto herramientas de software como lenguajes de programación como Python, para el análisis de diversos documentos de texto, enfocándose en la fase de preparación de datos.  Dichas funcionalidades se aplican a documentos que tratan sobre legislación ambiental en México, los cuales, aunque vienen en formato pdf, internamente presentan una estructura diversa, por lo que su análisis resulta más desafiante.  En este sentido, además de ejecutar tareas genéricas y específicas de preparación de los documentos, en este caso en los textos ambientales seleccionados como corpus, se ofrezcan datos cuantitativos acerca de la memoria y el tiempo transcurridos para las distintas configuraciones de tareas de preprocesamiento aplicadas, bajo una infraestructura de hardware determinadas.  El objetivo es medir los recursos computacionales como tiempo y memoria empleados para realizar tareas de preprocesamiento de documentos de texto, específicamente en la fase de preparación de datos mediante técnicas de minería de textos con la finalidad de caracterizar una dupla (una tarea y un recurso computacional) y de este modo, crear una instancia EC2 para aprovechar el cómputo en la nube. |