

Proyecto 1

La Escuela de Computación de la Universidad Central de Venezuela necesita un método eficiente para asignar la mención de las TEG que fueron publicadas antes de que existiera el sistema de ficheros y organización actual.

Para ello se han entregado datos de varias TEG que se encuentran en el sistema actual de la escuela. Dichas TEG vienen con la siguiente información:

- Título
- Resumen
- Palabras Claves
- Mención

Se desea que dado únicamente el título, resumen y palabras claves de una TEG se pueda predecir la mención a la que debe pertenecer dicho TEG.

En algunas menciones hay pocas apariciones de TEG y no proporcionan datos útiles para la minería, por lo que se limitará la minería a las siguientes menciones:

- Aplicaciones en Internet
- Base de Datos
- Inteligencia Artificial
- Sistemas de Información
- Tecnologías de Comunicación y Redes de Computadoras

Fase 1:

Para la primera fase se requiere que se haga todo el pre-procesamiento de los datos, esto incluye, limpieza, reducción y transformación de datos. Para esta fase se debe entregar:

- Un documento que explique todos los pasos tomados para preprocesar los datos.
- Si se utilizó alguna herramienta para el preprocesamiento, es necesario señalar en el documento que herramienta se utilizó.
- Si se implementó algún código para preprocesar debe incluirse el código fuente.
- Los archivos resultantes de hacer el preprocesamiento.

Fecha de Entrega: 12/10/2016

Fase 2:

Para la segunda etapa se debe haber completado la minería sobre los datos. Para esta fase se debe entregar:

- Un documento que explique los pasos realizados para hallar los patrones que permiten conseguir la mención a la que pertenece cada teg.
- Una aplicación que dado los datos de un nuevo teg pueda predecir a que mención pertenece.

Fecha de Entrega: 09/11/2016

Tip: Se recomienda implementar un algoritmo [Stemmer](#) para reducir las palabras a sus raíces y obtener información más útil.