



Resumo TFG

Identificación de TFG (para o profesorado):

BMGXITZ

Apelidos e Nome do/a Titor/a

Pavón Rial Reyes

Correo electrónico do/a Titor/a

pavon@uvigo.es

Apelidos e Nome do CoTitor/a (se procede)

Novo Lourés María

Apelidos e Nome do Alumno/a

Martin Perez Patricia

DNI do Alumno/a

Título do TFG

Pipe para pre-procesamiento de texto: análisis de sentimientos

Resumo

El marco de desarrollo de este proyecto está incluido en un proyecto de investigación desarrollado por el grupo SING (Sistemas Informáticos de Nueva Generación) inscrito en el registro de grupos de investigación de la Universidad de Vigo, cuyo objetivo principal es obtener mejoras en la eficacia de la clasificación de contenidos spam. Para conseguir esto, los contenidos o textos a clasificar como spam o legítimo deben ser pre-procesados. Para ello, se han desarrollado distintas funcionalidades o pipes que se encargan de llevar a cabo dicho pre-procesamiento y que consisten, por ejemplo, en obtener el idioma del texto, convertir los caracteres en minúscula o expandir las abreviaturas. De esta manera, se consigue tener un dataset que puede ser analizado por cualquier clasificador y determinar si es 'spam' o 'ham'.

Partiendo de esta base, la finalidad de este Trabajo Fin de Grado (TFG) es la realización de un nuevo pipe que sea capaz de extraer la polaridad del texto. De esta forma, el pipe recibirá un texto y se encargará de calcular la polaridad del mismo, determinando si el texto es de naturaleza negativa, positiva o neutra. Esta característica del texto será posteriormente analizada por el proyecto de investigación para ver su influencia en el clasificador de spam.

Para alcanzar el objetivo planteado en el TFG se realizará un

análisis de distintas técnicas empleadas en otros proyectos para obtener la polaridad de un texto, y posteriormente implementar el pipe correspondiente. Para ello se utilizará, posiblemente, un conjunto de lexicones de polaridades semánticas a nivel de palabras para diferentes idiomas. Además se llevará a cabo una aplicación que permita introducir un texto, aplicar el pipe desarrollado y mostrar el resultado obtenido. Toda el desarrollo se realizará en el lenguaje de programación Java empleando herramientas avanzadas de desarrollo y gestión de la configuración como Maven o Git. A mayores se empleará JSON para representar toda la información que sea posible de acuerdo con el software ya desarrollado por el grupo SING.

En cuanto a la metodología de desarrollo del proyecto, se ha decidido utilizar el Proceso Unificado el cuál facilitará el desarrollo iterativo e incremental completando una serie de tareas planificadas.

(A empresa debe ter convenio asinado en vigor coa Universidade de Vigo. Deberase entregar copia do nomeamento do/a titor/a pola empresa)

*** O/A profesor/a recibirá copia desta solicitude, e deberá dar o Visto e Prace do mesmo dende o formulario online dispoñible na web da ESEI.**

Código do TFG (para o alumno): El 19/20-77

**Introduce un correo electrónico válido
(a continuación recibirá un código
para a firma da solicitude)**

*** Unha vez enviada a solicitude recibirá por correo electrónico copia da mesma**