

## Resumo do Traballo de Fin de Grao

(Describe brevemente o traballo a desenvolver, xustificando o interese do mesmo, e indicando obxectivos, descrición técnica, proceso de desenvolvemento, e medios empregados. Engada tantas liñas como sexa necesario)

Las técnicas de detección de spam son capaces de detectar un porcentaje muy alto de contenido malicioso, pero el problema está lejos de resolverse, sobre todo por la cantidad de tráfico spam existente, no solo en el correo electrónico o SMS, sino también en redes sociales en Internet (como Facebook, Twitter, Instagram, etc.).

En el filtrado de spam basado en contenido, que podría ser utilizado para detectar spam en cualquiera de los servicios de comunicación disponible actualmente, se analiza el texto de los mensajes (e-mail, SMS, páginas web, comentario de YouTube, etc.) para determinar si el contenido es sospechoso. Dentro de las técnicas de análisis de texto, está teniendo gran auge en los últimos años el análisis de sentimiento (AS) de los mensajes para mejorar la detección de spam. El AS, extrayendo la polaridad de cada mensaje, proporciona información que permite ofrecer sistemas anti-spam más eficaces, mejorando los resultados de las técnicas de filtrado actuales.

En este contexto de filtrado de spam, el objetivo de este TFG es obtener la polaridad de los mensajes a partir de la información contenida en los mismos y ver cómo afecta a la clasificación spam / ham. Para ello, se hará uso de las interjecciones, emoticonos y emojis presentes en el texto.

Un segundo objetivo de este TFG es preprocesar los mensajes para extraer el significado de los emojis y emoticonos presentes en los mensajes para incorporarlo también en el texto a clasificar y enriquecer la información disponible para el clasificador.

El desarrollo de este proyecto se hará empleando la metodología ágil (por ejemplo, SCRUM), que posibilite la creación de un prototipo de una forma temprana al que, posteriormente, se le puedan incorporar funcionalidades adicionales.

Para el desarrollo se dará prioridad al uso de herramientas de código abierto (Maven, Git, etc,...). Concretamente, además de Git como repositorio de código y Maven para la gestión de la configuración del proyecto, se empleará el IDE Eclipse. Para el desarrollo se empleará un equipo portátil con procesador Intel I7 6700HQ, 8 GB de RAM, 1 Tb. de HDD y 256 Gb. de SSD. Los recursos computacionales de este equipo serán suficientes para el desarrollo que se pretende realizar.