Título: INGENIERÍA DE ATRIBUTOS Y MINERÍA DE DATOS PARA LA RECUPERACIÓN DE INFORMACIÓN CON ADVERSARIO

Nombre: PUERTAS SANZ, ENRIQUE

Universidad: Universidad Europea de Madrid

Departamento: INFORMÁTICA, AUTOMÁTICA Y COMUNICACIONES

Fecha de lectura: 19/12/2013

Programa de doctorado: DOCTORADO EN TECNOLOGÍAS DE LA INFORMACIÓN APLICADAS

Dirección:

> **Director:** JOSÉ MARÍA GÓMEZ HIDALGO

Tribunal:

> **presidente:** MARÍA TERESA VILLALBA DE BENITO

> **secretario:** MARIA CRUZ GAYA LOPEZ

> **vocal:** Carlos Laorden Gómez

> **vocal:** Luis Martín Martín

> **vocal:** JUAN MARTÍNEZ ROMO

Descriptores:

> INTELIGENCIA ARTIFICIAL

> LENGUAJES DE PROGRAMACION

> LINGUISTICA COMPUTACIONAL

> DOCUMENTACION AUTOMATIZADA

El fichero de tesis ya ha sido incorporado al sistema

Localización: UNIVERSIDAD EUROPEA DE MADRID

Resumen: El creciente uso de Internet ha venido acompañado de numerosas ventajas, pero también de oportunidades para el fraude. Un buen ejemplo de este tipo de abuso lo encontramos en el correo electrónico, una herramienta con indudable valor para la comunicación de las personas, pero que tiene el inconveniente del correo no solicitado (spam). Otros abusos son, por ejemplo, la descarga de páginas web inapropiadas (e.g. pornográficas) en el puesto de trabajo, o el spam enviado a dispositivos móviles. Debido a la naturaleza de índole textual que se maneja en ese tipo de escenarios, éstos han sido abordados normalmente por medio de técnicas de minería de texto, es decir, de descubrimiento de conocimiento en bases de datos textuales. Sin embargo, ese tipo de abusos tienen elemento común que hace que las tareas de minería de texto tradicionales no funcionen correctamente: En todas ellas existe un adversario que intenta degradar la eficiencia de los categorizadores de texto generados por técnicas de aprendizaje automático. En estos casos se habla de tareas de clasificación o categorización (de texto) con adversario, en el que los sistemas de análisis y aprendizaje deben tener presente la existencia de un adversario (por ejemplo, el spammer) cuyo objetivo es degradar la

efectividad de los sistemas de clasificación construidos con estas técnicas. En esta Tesis, las dos contribuciones fundamentales del trabajo son la aplicación de técnicas de ingeniería de atributos y el desarrollo de un método específico de evaluación, más adecuado que los precedentes, para este tipo de problemas con adversario. Éste método de evaluación que hemos propuesto en esta investigación se ha convertido en un estándar en el campo científico de la seguridad, y se ha utilizado en competiciones científicas del más alto nivel, como las Conferencias TREC (Text REtrieval Conferences), para la evaluación de sistemas de filtrado de correo basura. Más concretamente, en esta Tesis hemos demostrado que es posible tratar de una manera unificada el proceso más sensible en la Categorización de Texto con Adversario, que es la representación de los textos, usando técnicas de ingeniería del Lenguaje Natural, y realizar una evaluación homogénea para diversas tareas a pesar de los distintos costes, variables, y de los distintas asimetrías en la distribución de las clases.