

---

## Trabajo Fin de Máster: Título del Trabajo

---



### Trabajo Fin de Máster

**Nombre del Alumn@**

Trabajo de investigación para el

Máster en Lenguajes y Sistemas Informáticos

Universidad Nacional de Educación a Distancia

Dirigido por el

**Prof. Dr. D. Nombre Direct@r**

Marzo 2015



# Agradecimientos

Agradecimientos si procede.



# Resumen

Breve resumen del trabajo realizado y de los objetivos conseguidos.



# Abstract

Breve resumen del trabajo realizado y de los objetivos conseguidos en inglés.





# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Propuesta y objetivos . . . . .	1
1.3. Estructura del documento . . . . .	1
<b>2. Estado del arte</b>	<b>3</b>
2.1. Detección de lenguaje o discurso del odio ( <i>hate speech de- tection</i> ) . . . . .	3
2.2. Ejemplo sección . . . . .	5
2.2.1. Ejemplo subsección . . . . .	5
<b>3. Sistema/Método/Caso de Estudio propuesto</b>	<b>7</b>
3.1. Ejemplo sección . . . . .	7
3.1.1. Ejemplo subsección . . . . .	7
<b>4. Evaluación</b>	<b>9</b>
4.1. Metodología de evaluación . . . . .	9
4.2. Métricas de evaluación . . . . .	9
4.3. Colecciones de evaluación . . . . .	9
4.4. Resultados . . . . .	9
<b>5. Discusión</b>	<b>11</b>
5.1. Ejemplo sección . . . . .	11
5.1.1. Ejemplo subsección . . . . .	11
<b>6. Conclusiones y trabajo futuro</b>	<b>13</b>
6.1. Conclusiones . . . . .	13
6.2. Trabajo futuro . . . . .	13

<b>Bibliografía</b>	<b>15</b>
<b>A. Publicaciones</b>	<b>17</b>

# Índice de Figuras



# Índice de Tablas



# Capítulo 1

## Introducción

### 1.1. Motivación

Motivación del trabajo a realizar.

El primer párrafo de cada sección no se indenta, los siguientes sí. Las referencias a un artículo del estado del arte se ponen así: El trabajo bla bla.

La presente plantilla es meramente orientativa. Tanto los capítulos como las secciones de cada capítulo son simplemente una guía para ayudar al alumno en el proceso de escritura del trabajo final del máster, pero en ningún caso supone que los trabajos tengan que tener forzosamente esta estructura. La estructura del trabajo debe ser acordada por el alumno y su director/res, así como adecuarse a la temática abordada.

### 1.2. Propuesta y objetivos

Que se ha realizado en el trabajo de fin de master, cuales eran los objetivos y breve resumen de los resultados obtenidos. Texto de prueba 3.

### 1.3. Estructura del documento

Breve descripción de los capítulos del trabajo.

**Capítulo 1. Introducción.** Este capítulo introduce los principales motivos que han llevado a la realización de este trabajo, así como la problemática y el estado actual de la disciplina. Por último, se presentan las diferentes contribuciones del trabajo realizado.

**Capítulo 2. Estado del arte.** Este capítulo describe en mayor detalle la disciplina que nos ocupa, presentando su origen y su historia hasta el presente. Se muestran las técnicas actuales más utilizadas para resolver las tareas más relevantes del tema abordado, así como sus debilidades.

**Capítulo 3. Sistema/Método/Caso de Estudio propuesto.** En este capítulo se describe en profundidad el sistema/método o caso de estudio propuesto.

**Capítulo 4. Evaluación.** Este capítulo describe la metodología utilizada para evaluar la propuesta realizada, a la vez que presenta los resultados obtenidos al evaluar el método propuesto en diferentes tareas y sobre colecciones de evaluación de distintos dominios.

**Capítulo 5. Discusión.** Este capítulo analiza y discute en profundidad los resultados obtenidos en la evaluación presentada en el capítulo anterior.

**Capítulo 6. Conclusiones y trabajo futuro.** Este capítulo recopila las diferentes conclusiones extraídas del trabajo realizado, y propone algunas líneas de trabajo futuro.



## Capítulo 2

# Estado del arte

El presente capítulo tiene como objetivo presentar al lector la detección del lenguaje machista en redes sociales. Para ello, se realizará una revisión de los trabajos más relevantes en la tarea de detección de lenguaje abusivo y machista, en los que se analizarán los orígenes de esta tarea, las soluciones técnicas y las aportaciones más relevantes.

### 2.1. Detección de lenguaje o discurso del odio (*hate speech detection*)

La detección del lenguaje machista o sexista está muy relacionada con la detección del lenguaje o discurso del odio en redes sociales. Existen numerosos trabajos donde se intenta detectar distintos tipos de lenguaje del odio, entre ellos el sexismo (HAJIME WATANABE, 2018; Zeerak Waseem, 2016; Georgios K. Pitsilis y Langseth, 2018; Pinkesh Badjatiya, 2017; Steven Zimmerman, 2018; Park y Fung, 2017; Waseem, 2016). El lenguaje del odio se refiere al uso de lenguaje agresivo, violento u ofensivo hacia un grupo específico de personas que comparten una propiedad en común, sea esta propiedad su género, su raza, sus creencias o su religión (Thomas Davidson, 2017). Atendiendo a esta definición, se puede considerar la detección del machismo como un caso particular del discurso del odio. Por ello, es muy interesante realizar una evaluación de los trabajos realizados en esta línea de investigación.

La detección del lenguaje del odio es una línea de investigación muy actual, el primer estudio evaluado data del año 2012 (Guang Xiang, 2012). En

este artículo se emplea un modelo de detección de temas o categorías (*topic modelling*) que explota la concurrencia de palabras para la creación de atributos o *features* que alimentarán un algoritmo de clasificación de aprendizaje de máquina o *machine learning*. En la mayoría de trabajos previos se empleaban soluciones basadas en patrones para la clasificación de tweets. De este modo, este artículo supone un paso muy importante hacia la automatización y a los sistemas basados en algoritmos de *machine learning*. Además, durante la etapa anterior a este artículo, el uso de expresiones coloquiales y soeces en redes sociales hace difícil establecer las fronteras entre el uso de lenguaje ofensivo que no tiene como objetivo despreciar a ningún grupo de personas y el lenguaje del odio (Thomas Davidson, 2017) utilizando patrones extraídos de la utilización del lenguaje.

Durante los últimos tres años, se han sucedido los artículos en la temática y ha aumentado considerablemente la producción científica en este campo. En (Zeeraak Waseem, 2016) se aporta el primer corpus de referencia anotados que se utilizará posteriormente en (Waseem, 2016; Georgios K. Pitsilis y Langseth, 2018; Pinkesh Badjatiya, 2017; Steven Zimmerman, 2018; Park y Fung, 2017). Está compuesto por 16.000 *tweets* etiquetados para mensajes sexistas, racistas o sin contenido ofensivo. En este primer trabajo, se sientan las bases de las soluciones aplicadas en el resto de artículos, se utilizan atributos como los *unigramas*, *bigramas*, *trigramas* y *cuatri-gramas* y un algoritmo de regresión logística para la clasificación.

En el artículo desarrollado por el mismo autor (Waseem, 2016) se propone una solución similar pero se amplía el corpus en 4033 *tweets* y se utiliza una plataforma de *crowdsourcing* para anotar los mensajes. Achacan el empeoramiento de los resultados al posible sesgo que se produce en (Zeeraak Waseem, 2016) ya que los *tweets* solo fueron etiquetados por los autores únicamente.

En el resto de artículos que evalúan su propuesta utilizando el corpus desarrollado por (Waseem, 2016), se utilizan redes neuronales para la tarea de clasificación y, en algunos, en la etapa de preprocesamiento. En la solución propuesta por (Steven Zimmerman, 2018) se aplican redes neuronales convolucionales (*CNN*, *Convolutional Neural Network*) para codificar el texto y extraer los atributos que se utilizarán para el clasificador final, basado también en CNNs. Esta técnica permite tener en cuenta la posición de la palabra (su contexto) para extraer los atributos de cada *tweet*. Esta

misma idea junto con el uso de redes neuronales recurrentes (*RNN*, *Recurrent Neural Network*) se utiliza en (Pinkesh Badjatiya, 2017) para obtener los atributos en la etapa de procesamiento. En ambos artículos se consiguen mejorar los resultados alcanzados por (Waseem, 2016).

En (Georgios K. Pitsilis y Langseth, 2018) se propone un modelo basado en RNNs para abordar el problema. Además se explora la idea de utilizar atributos como la tendencia al racismo o sexismo utilizando el historial de los usuarios. Se demuestra como el uso de este tipo de atributos mejora notablemente los resultados.

En todos los artículos revisados anteriormente, se trata el problema como una clasificación múltiple donde el texto se puede clasificar según las etiquetas racismo, sexismo o ninguno. Sin embargo, se podría resolver el problema con un doble clasificador, el primero clasifica si el texto contiene lenguaje abusivo o no y el segundo realizaría la tarea de clasificar en contenido sexista o racista (Park y Fung, 2017).

## 2.2. Ejemplo sección

### 2.2.1. Ejemplo subsección



## Capítulo 3

# Sistema/Método/Caso de Estudio propuesto

En este capítulo se describe en profundidad el sistema/método o caso de estudio propuesto. La organización de este capítulo dependerá sustancialmente del trabajo abordado.

### 3.1. Ejemplo sección

#### 3.1.1. Ejemplo subsección



## Capítulo 4

# Evaluación

Este capítulo describe la metodología utilizada para evaluar el sistema/método o caso de estudio propuesto, a la vez que presenta los resultados obtenidos en la evaluación de las diferentes tareas y sobre colecciones de evaluación de distintos dominios. Algunos ejemplos de secciones pueden ser estos:

### 4.1. Metodología de evaluación

### 4.2. Métricas de evaluación

### 4.3. Colecciones de evaluación

### 4.4. Resultados





# Capítulo 5

## Discusión

Este capítulo analiza y discute en profundidad los resultados obtenidos en la evaluación presentada en el capítulo anterior. La estructura de este capítulo dependerá del tema del trabajo y de la estructura del capítulo anterior.

### 5.1. Ejemplo sección

#### 5.1.1. Ejemplo subsección



## Capítulo 6

# Conclusiones y trabajo futuro

Este capítulo recopila las diferentes conclusiones extraídas del trabajo realizado, y propone algunas líneas de trabajo futuro. Las siguientes secciones son las que suelen contener este tipo de capítulos, aunque pueden variar dependiendo del tema del trabajo.

### 6.1. Conclusiones

### 6.2. Trabajo futuro



# Bibliografía

## Bibliografía

- [Georgios K. Pitsilis y Langseth2018] Georgios K. Pitsilis, Heri Ramampiaro y Helge Langseth. 2018. Detecting oensive language in tweets using deep learning. En *Department of Computer Science Norwegian University of Science and Technology*.
- [Guang Xiang2012] Guang Xiang, Bin Fan, Ling Wang Jason I. Hong Carolyn P. Rose. 2012. Detecting offensive tweets via topical feature discovery over a large scale twitter corpus. En *CIKM, 2012, Maui, HI, USA*.
- [HAJIME WATANABE2018] HAJIME WATANABE, MONDHER BOUAZIZI, TOMOAKI OHTSUKI. 2018. Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. En *2018 IEEE*.
- [Park y Fung2017] Park, Ji Ho y Pascale Fung. 2017. One-step and two-step classification for abusive language detection on twitter. En *Proceedings of the First Workshop on Abusive Language Online*.
- [Pinkesh Badjatiya2017] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta Vasudeva Varma. 2017. Deep learning for hate speech detection in tweets. En *2017 International World Wide Web Conference Committee (IW3C2)*.
- [Steven Zimmerman2018] Steven Zimmerman, Chris Fox, Udo Kruschwitz. 2018. Improving hate speech detection with deep learning ensembles. En *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

- [Thomas Davidson2017] Thomas Davidson, Dana Warmusley, Michael Macy Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. En *ICWSM 2017*.
- [Waseem2016] Waseem, Zeerak. 2016. Are you a racist or am i seeing things? annotator influence on hate speech detection on twitter. En *Proceedings of 2016 EMNLP Workshop on Natural Language Processing and Computational Science*.
- [Zeerak Waseem2016] Zeerak Waseem, Dirk Hovy. 2016. Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. En *Proceedings of NAACL-HLT 2016*, páginas 33–41.

## Apéndice A

# Publicaciones

Publicaciones derivadas del trabajo realizado.

**Todo list**