

# Tabla de contenidos

Agradecimientos	ii
Resumen	iii
Resum	iv
Abstract	v
Tabla de contenidos	vi
Lista de Figuras	xi
Lista de Tablas	xiii
<b>1 Introducción</b>	<b>1</b>
1.1 Motivación y objetivos	2
1.1.1 Cuestiones de investigación	3
1.1.2 Objetivos	3
1.2 Aportaciones de la tesis	4
1.3 Organización de la tesis	4
<b>2 Identificación de edad y sexo</b>	<b>7</b>
2.1 Trabajos previos	7
2.2 Lenguaje y social media	14
2.2.1 Metodología	14
2.2.2 Riqueza léxica	15
2.2.3 Distribución gramatical	16
2.2.4 Uso del lenguaje por sexo	19
2.2.5 Discusión	20
2.3 Conclusiones	20
<b>3 Author profiling en el PAN</b>	<b>21</b>
3.1 Author profiling en PAN 2013	22
3.1.1 Corpus	22
3.1.2 Aproximaciones	23
3.1.3 Resultados	24
3.1.4 Discusión	26
3.2 Author profiling en PAN 2014	26
3.2.1 Corpus	26
3.2.2 Aproximaciones	30
3.2.3 Resultados	31
3.2.4 Resultados entre años	35
3.2.5 Discusión	36
3.3 Author profiling en PAN 2015	36
3.3.1 Corpus	36
3.3.2 Aproximaciones	37
3.3.3 Resultados	39

3.3.4	Discusión	42
3.4	Conclusiones	43
<b>4</b>	<b>Identificación de emociones en medios sociales</b>	<b>45</b>
4.1	Trabajos previos	45
4.1.1	Generación de recursos afectivos	45
4.1.2	Métodos de procesamiento afectivo	46
4.2	Emociones y tendencias en Twitter	48
4.2.1	Metodología	48
4.2.2	Resultados	49
4.2.3	Discusión	51
4.3	Emociones y sexo en Facebook	51
4.3.1	Características de estilo	52
4.3.2	Metodología	53
4.3.3	Resultados	56
4.3.4	Discusión	58
4.4	Emociones, sexo e ironía en Facebook	59
4.4.1	Metodología	59
4.4.2	Emociones	60
4.4.3	Ironía	62
4.4.4	Discusión	64
4.5	Emociones, sexo y edad en PAN	64
4.5.1	Metodología	64
4.5.2	Resultados	65
4.5.3	Discusión	66
4.6	Conclusiones	66
<b>5</b>	<b>EmoGraph: Una aproximación basada en grafos</b>	<b>67</b>
5.1	Grafos en procesamiento del lenguaje natural	68
5.2	EmoGraph	68
5.2.1	Construcción y enriquecimiento del grafo	68
5.2.2	Características	72
5.2.3	Metodología	74
5.2.4	Resultados	75
5.2.5	Discusión	76
5.3	Robustez ante idiomas y social media	86
5.3.1	Metodología	86
5.3.2	Resultados	87
5.3.3	Contribución de EmoGraph	88
5.3.4	Discusión	88
5.4	Conclusiones	89
<b>6</b>	<b>Identificación del lenguaje nativo y de las variedades del lenguaje</b>	<b>91</b>
6.1	Identificación del lenguaje nativo	92
6.1.1	Corpus	92
6.1.2	Aproximaciones	93
6.2	Identificación de lenguas similares y variedades del lenguaje	94
6.2.1	Corpus	94
6.2.2	Aproximaciones	95
6.3	Conclusiones	96
<b>7</b>	<b>Aproximaciones para la identificación de variedades del lenguaje</b>	<b>97</b>
7.1	HispaBlogs	98
7.2	Representaciones distribuidas	99
7.2.1	Modelo continuo de Skip-gramas	99
7.2.2	Metodología	100
7.2.3	Resultados	103
7.2.4	Análisis del error	103

---

7.2.5	Discusión	104
7.3	Representación de baja dimensionalidad	105
7.3.1	Esquema de representación	105
7.3.2	Metodología	107
7.3.3	Resultados	107
7.3.4	Discusión	108
7.4	Identificación de idiomas similares	115
7.4.1	Descripción de la tarea	115
7.4.2	Representaciones distribuidas	117
7.4.3	Representación de baja dimensionalidad	119
7.4.4	Comparativa entre representaciones	121
7.5	Conclusiones	122
<b>8</b>	<b>Conclusiones y trabajo futuro</b>	<b>125</b>
8.1	Contribuciones	127
8.2	Trabajo futuro	127
8.3	Publicaciones	129
8.4	Impacto en medios	131
8.5	Difusión en medios sociales	132
	<b>Bibliografía</b>	<b>135</b>
	<b>Apéndice I. Author Profiling en PAN 2014</b>	<b>151</b>
A	Significación estadística pareada de sistemas	151
B	Distancias en la identificación de edad	160
	<b>Apéndice II. Author Profiling en PAN 2015</b>	<b>163</b>
C	Significación estadística pareada entre sistemas	163
D	Significación estadística pareada de sistemas por idioma	165