

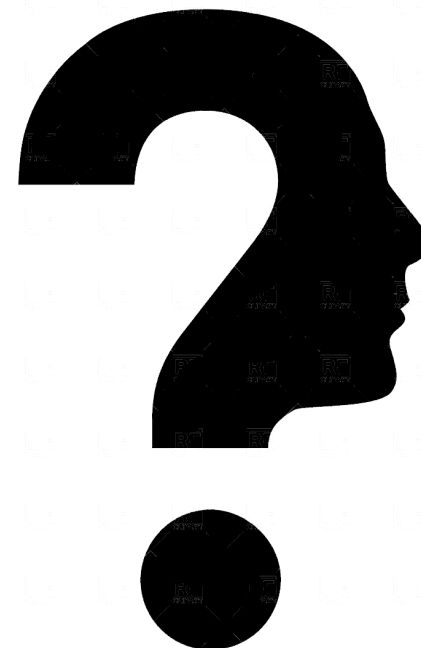
# **Perfiles de Autor: Identificación de Género y Edad**

**Francisco Manuel Rangel Pardo**  
**Dirige Dr. Paolo Rosso**

## Introducción

► La tarea de Author Profiling pretende obtener, a partir de textos anónimos, información de su autor como por ejemplo:

- Género
- Edad
- Idioma nativo (var. regional)
- Emociones
- Rasgos de personalidad
- ...
- Robots vs. humanos
- Ideología política
- Creencias religiosas
- Profesión
- ...



## Motivación

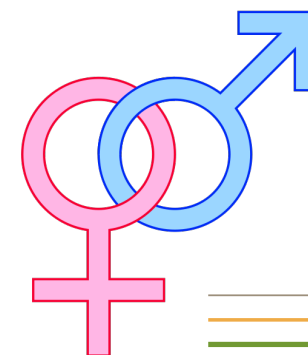
- Punto de vista forense -> Lenguaje como evidencia
- Punto de vista de la seguridad -> Perfil de posibles delincuentes
- Punto de vista del marketing -> Segmentación de usuarios

## Objetivo

- Modelar cómo las personas expresan emociones cuando escriben en las redes sociales y utilizar esta información para identificar su género, edad, idioma nativo o variedad regional, o sus rasgos de personalidad

## Estado actual

- Organización de la tarea Author Profiling en PAN@CLEF (2013, 2014, ...)
- Aproximación basada en teoría de grafos y aprendizaje automático para modelar el impacto de las emociones en los perfiles de autor -> identificación de edad y género

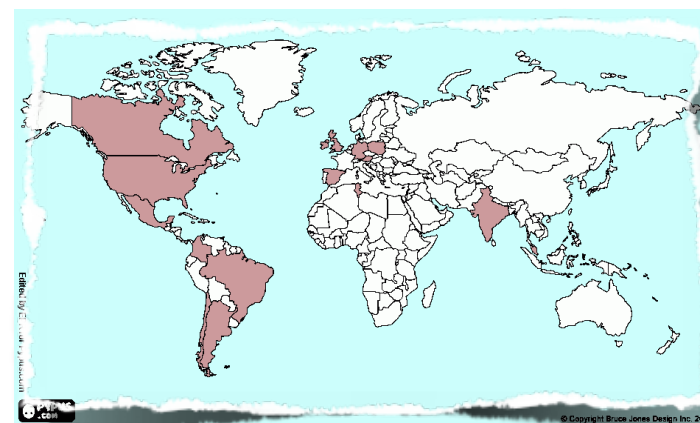


## Author Profiling en PAN@CLEF 2013

- Identificación de edad (10s, 20s, 30s) y género en redes sociales en inglés y español

IDIOMA	NÚMERO DE AUTORES		
	ENTRENAMIENTO	EARLY BIRDS	TEST
<b>EN</b>	<b>236.600</b>	<b>21.200</b>	<b>25.440</b>
<b>ES</b>	<b>75.900</b>	<b>6.800</b>	<b>8.160</b>

- 66 equipos registrados
- 21 participantes
- 16 países
- 18 artículos (8 long / 10 short)



## Features

### Author Profiling en PAN@CLEF 2013

- ▶ Frecuencias de signos de puntuación, letras mayúsculas, comillas...
- ▶ Categorías gramaticales
- ▶ Tags HTML como <image> y <a>
- ▶ Medidas varias de legibilidad
- ▶ Emoticonos
- ▶ LSA, BoW, TF-IDF, n-gramms
- ▶ Entidades nombradas, palabras de sentimiento, palabras de emociones
- ▶ Slang, contracciones, palabras alargadas (holaaaa...)
  
- ▶ Colocaciones -> GANADOR EN INGLÉS
- ▶ Representación de segundo orden basada en la relación entre documentos y perfiles -> GANADOR DE LA TAREA

## Author Profiling en PAN@CLEF 2014



### Colaboración PAN-Replab

Dos perspectivas complementarias del **Author Profiling**

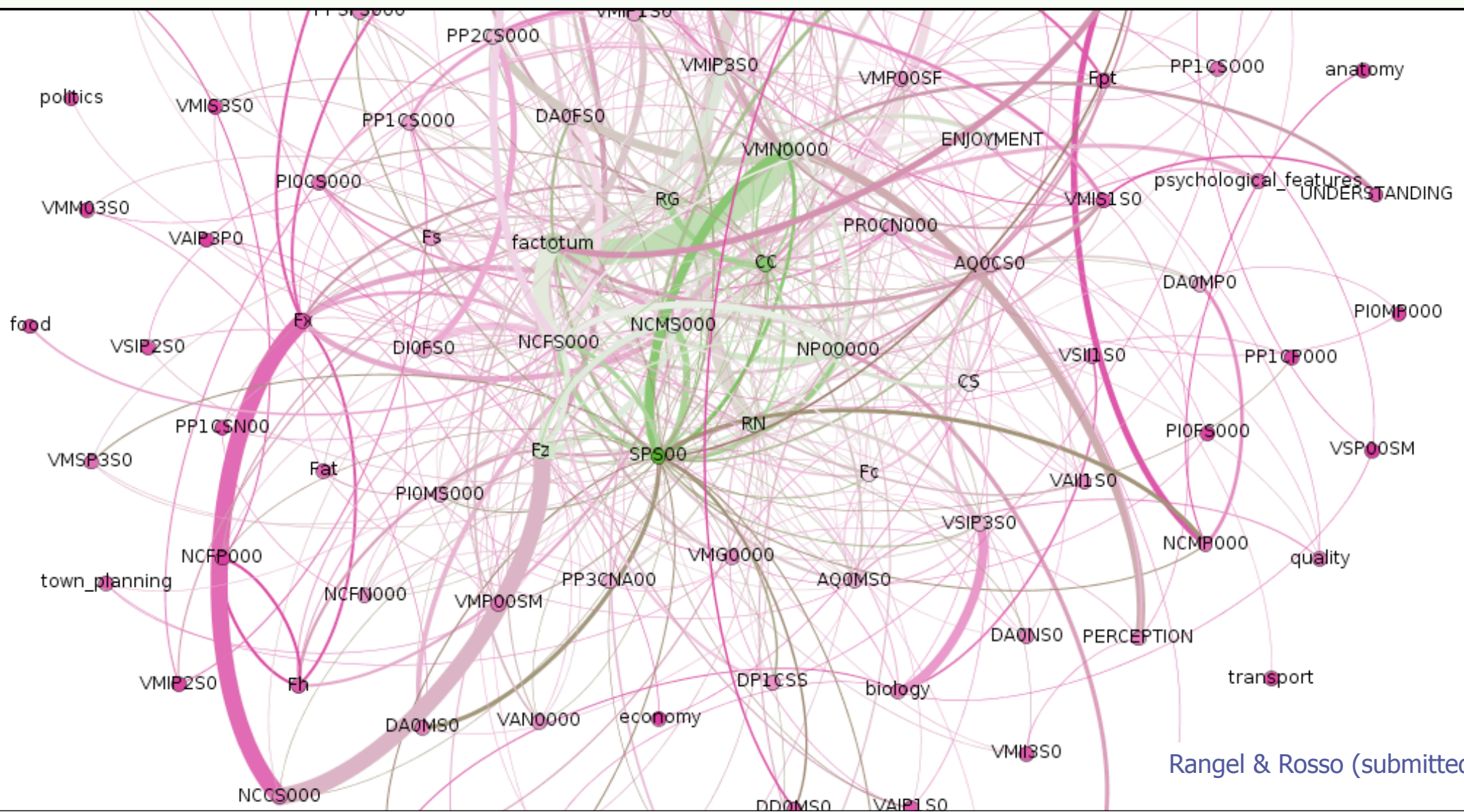
**Máquinas virtuales de PAN** para participantes de **RepLab TIRA platform @ Weimar**

- Participantes con la misma **capacidad computacional**
- Incrementa el **engagement** de los participantes
- Mejora **Sostenibilidad, Replicabilidad y Reproducibilidad**
- Permite **evaluaciones cross-year**

### Nuevos datasets

- PAN-AP13 -> Social Media
- TripAdvisor (EN)
- Blogs
- Twitter (con Replab)

# Impacto de las Emociones en los Perfiles de Autor: Una aproximación basada en grafos con etiquetas emocionales



Rangel & Rosso (submitted, 2014)



## Resultados con dataset PAN-AP13

Ranking	Team	Accuracy
1	<b>Rangel-EG</b>	0.6624
2	Pastor	0.6558
3	Santosh	0.6430
4	<b>Rangel-S</b>	0.6350
5	Haro	0.6219
6	Flekova	0.5966
...	...	
21	Baseline	0.3333
...	...	
23	Mechti	0.0512

Identificación de Edad

Ranking	Team	Accuracy
1	Santosh	0.6473
2	<b>Rangel-EG</b>	0.6365
3	Pastor	0.6299
4	Haro	0.6165
5	Ladra	0.6138
...	...	
8	<b>Rangel-S</b>	0.5713
...	...	
18	Baseline	0.5000
...	...	
23	Gillam	0.4784

Identificación de Género



## Palabras más usadas por género

autoritas consulting, s.a



Females

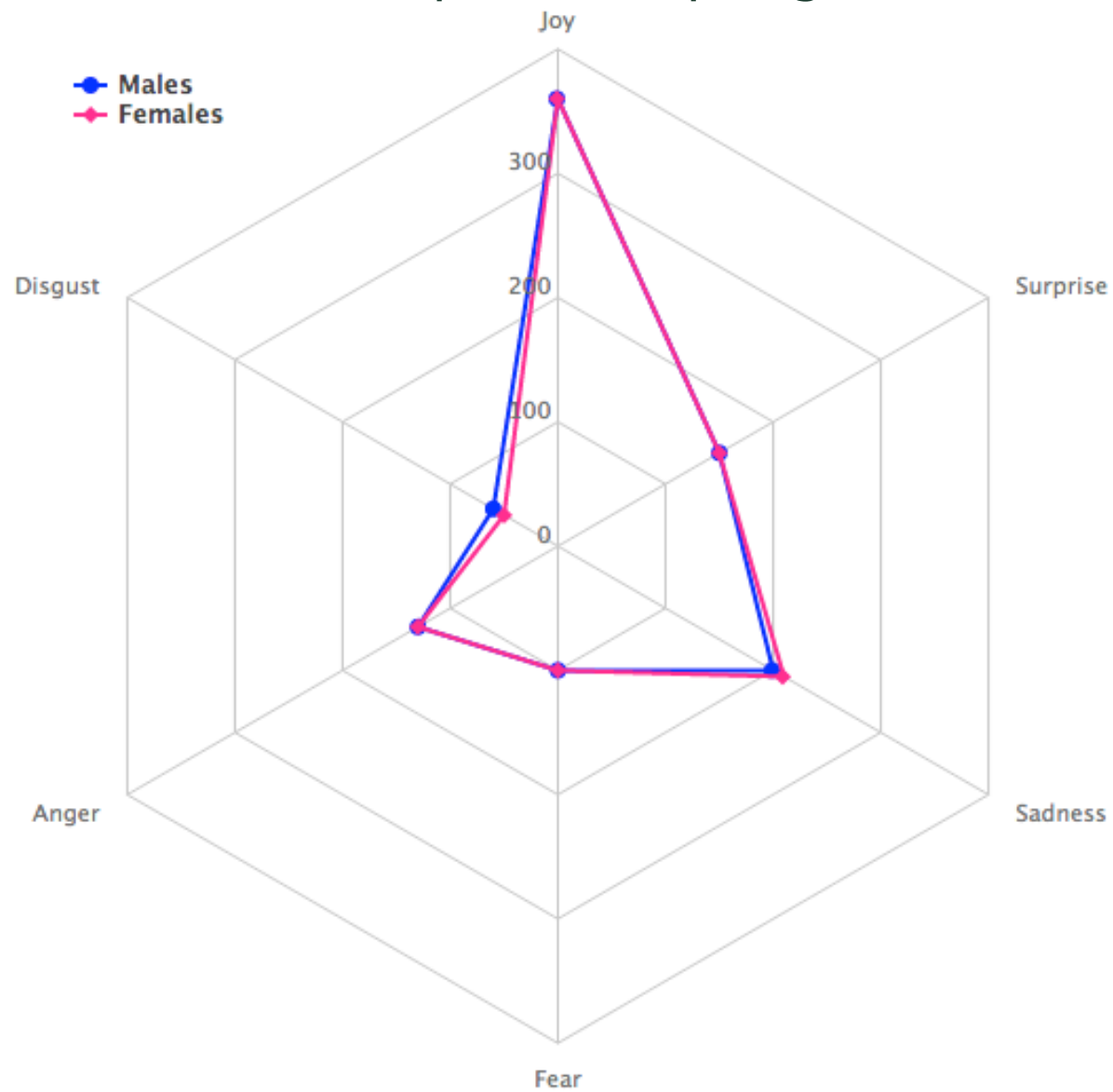
Males



Rangel & Rosso (submitted, 2014)

## Emociones más expresadas por género

autoritas consulting, s.a



Rangel & Rosso (submitted, 2014)

## Tipos de verbos por género\*

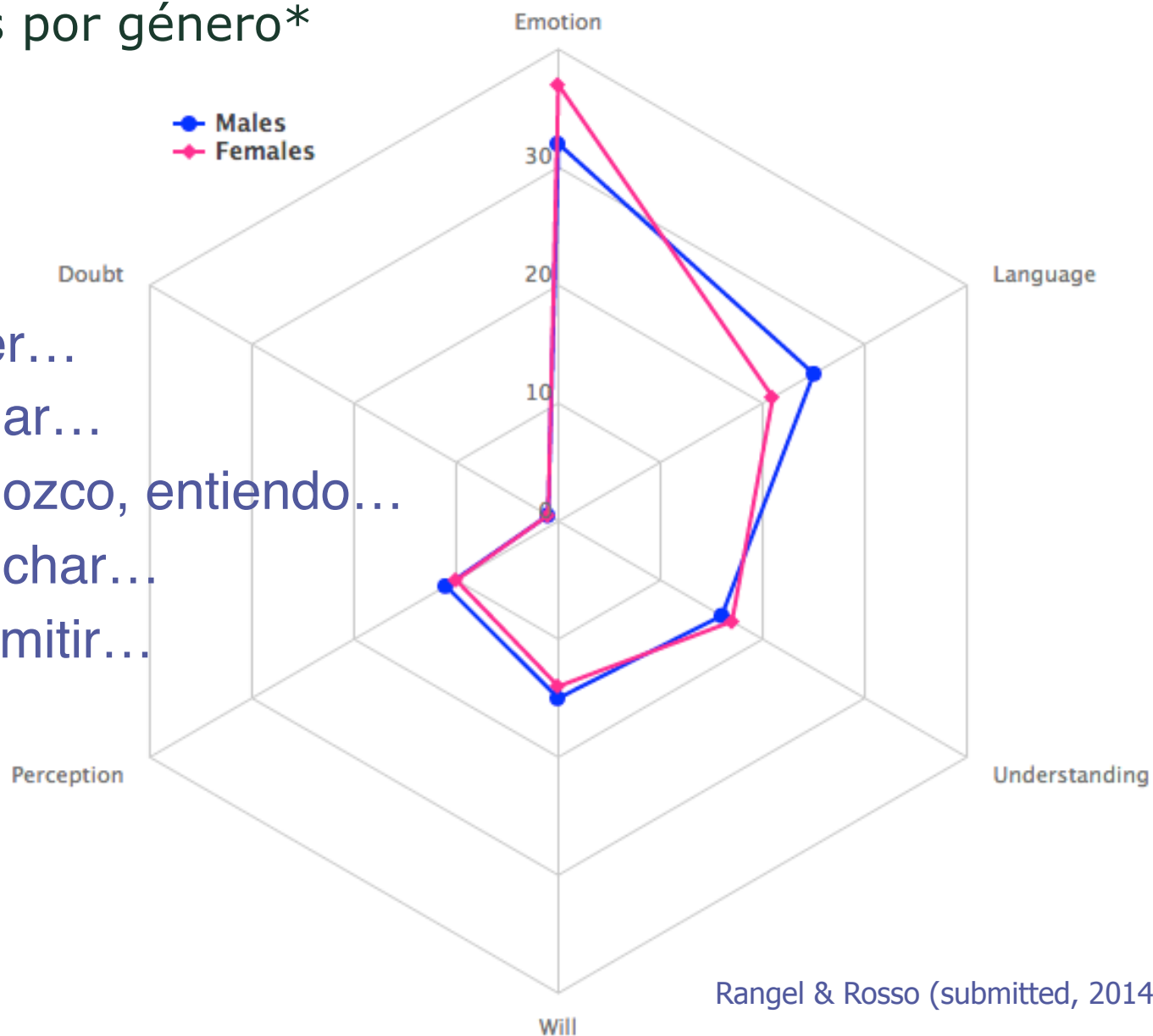
**Emoción:** sentir, amar, querer...

**Lenguaje:** decir, contar, hablar...

**Entendimiento:** pienso, conozco, entiendo...

**Percepción:** ver, sentir, escuchar...

**Permiso:** deber, prohibir, permitir...



\*Levin (1993)

Rangel & Rosso (submitted, 2014)

## Conclusiones

- ▶ Las emociones expresadas en los textos, modeladas mediante grafos, permiten mejorar la identificación de género y edad en la tarea de Author Profiling

## Trabajo futuro

- ▶ Extender el trabajo al inglés
- ▶ Aplicar la propuesta a la identificación de rasgos de personalidad
  - ▶ Introducción de la tarea en PAN 2015



**Francisco Rangel**  
Autoritas Consulting /  
Universitat Politècnica de València  
[francisco.rangel@autoritas.es](mailto:francisco.rangel@autoritas.es)  
[@kicorangel](#)



**Paolo Rosso**  
Universitat Politècnica de València  
[prossor@dsic.upv.es](mailto:prossor@dsic.upv.es)

*¡Muchas gracias  
por vuestra  
atención!*