

# Author Profiling at PAN: from Age and Gender Identification to Language Variety Identification (invited talk)

**Paolo Rosso**

Natural Language Engineering Lab  
PRHLT Research Center  
Universitat Politècnica de València  
Valencia, Spain  
prossso@dsic.upv.es

## 1 Abstract

Author profiling is the study of how language is shared by people, a problem of growing importance in applications dealing with security, in order to understand who could be behind an anonymous threat message, and marketing, where companies may be interested in knowing the demographics of people that in online reviews liked or disliked their products. In this talk we will give an overview of the PAN<sup>1</sup> shared tasks that since 2013 have been organised at CLEF and FIRE evaluation forums, mainly on age and gender identification in social media, although also personality recognition in Twitter as well as in code sources was also addressed.

In 2017 the PAN author profiling shared task addresses jointly gender and language variety identification in Twitter where tweets have been annotated with authors' gender and their specific variation of their native language: English (Australia, Canada, Great Britain, Ireland, New Zealand, United States), Spanish (Argentina, Chile, Colombia, Mexico, Peru, Spain, Venezuela), Portuguese (Brazil, Portugal), and Arabic (Egypt, Gulf, Levantine, Maghrebi).

## 2 Biography

Paolo Rosso<sup>2</sup> is an associate professor of computer science at the Technical University of Valencia, Spain where he leads the NLE lab of the PRHLT research center. His research interests include author profiling and irony detection in social media, opinion spam detection, as well as text reuse and plagiarism detection. Since 2009 he has been involved in the organisation of PAN benchmark activities, since 2010 and 2011 in the framework of CLEF and FIRE evaluation forums, on plagiarism / text reuse detection and author profiling. He has been also co-organiser of the shared task on Sentiment analysis of figurative language in Twitter at SemEval-2015. Paolo Rosso has been PI in several national and international research projects. He is the co-author of 50+ articles in international journals and 400+ articles in conferences and workshops. He has been advisor of 17 PhD students. He has been chair of \*SEM-2015, and organisation chair of CLEF-2013 and also of EACL-2017 both here in Valencia...

---

<sup>1</sup><http://pan.webis.de/>

---

<sup>2</sup><http://www.dsic.upv.es/~prossso/>