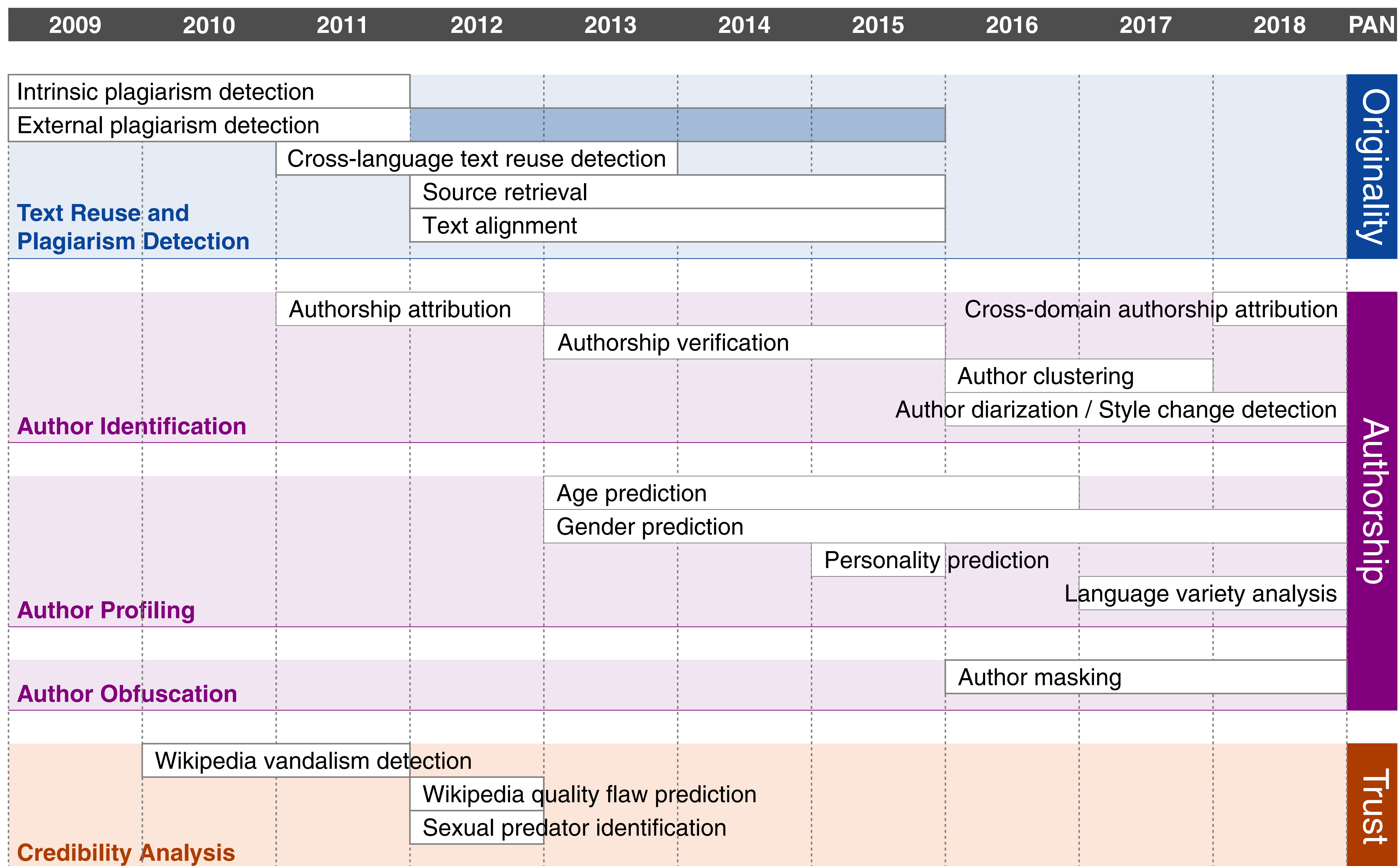


A Decade of Shared Tasks in Digital Text Forensics at PAN

Martin Potthast¹ Paolo Rosso² Efstathios Stamatatos³ Benno Stein⁴

¹ University of Leipzig, Germany
² Universitat Politècnica de València, Spain
³ University of the Aegean, Greece
⁴ Bauhaus-Universität Weimar, Germany



PAN Shared Tasks

Digital text forensics aims at examining the **originality** and **credibility** of information in electronic documents and, in this regard, to extract and analyze information about the **authors** of these documents. The research field has been substantially developed during the last decade. PAN is a series of shared tasks that started in 2009 and significantly contributed to attract the attention of the research community in well-defined digital text forensics tasks. Several benchmark datasets have been developed to assess the state-of-the-art performance in a wide range of tasks. The evolution of shared tasks so far is depicted in the figure.

PAN 2019

- **Bots and gender profiling:** Given a Twitter feed, determine whether its author is a bot or a human. In case of human, identify her/his gender.
- **Celebrity profiling:** Given the English Social Media feed of a celebrity, determine their degree of fame, occupation, age, and gender.
- **Cross-domain authorship attribution:** Given a fanfiction text, determine its author among a list of candidates.
- **Style change detection:** Given a document, determine whether it is written by a single or multiple authors. In the latter case, determine the number of involved collaborators.