

CLEF 2014



Martin Potthast, Benno Stein,
Efsthathios Stamatatos, Paolo Rosso,
Patrick Joula, Walter Daelemans,
Francisco Rangel, etc. (many many
more people, mainly from Weimar)

CLEF Labs @ ECIR 2014
Monday, April 14

Francisco Rangel
Autoritas / Universitat Politècnica de València

What is PAN?

10th workshop/task/lab on

**Uncovering Plagiarism, Authorship,
and Social Software Misuse**

PAN: Workshops



Bauhaus-Universität Weimar:

Benno Stein's Webis group (Martin Potthast etc.)

<http://www.webis.de/research/events/pan-07>



Plagiarism detection

Wikipedia vandalism detection

<http://www.webis.de/research/events/pan-10>



+ **University of the Aegean**

(Efsthathios Stamatatos) and

Bar-Ilan University (Moshe Koppel)

<http://www.webis.de/research/events/pan-08>



Plagiarism detection

Wikipedia vandalism detection

Authorship identification

+ **Duquesne University** (Patrick Juola) and

Illinois Institute of Technology (Shlomo Argamon)

<http://www.webis.de/research/events/pan-11>



Plagiarism detection (source retrieval / text alignment)

Authorship identification

Author profiling (age and gender identification in social media; in English and Spanish)

+ **Autoritas Consulting** (Francisco Rangel)

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



Task on Plagiarism detection (external / intrinsic)

+ **Universitat Politècnica de València** (Paolo Rosso, Alberto Barrón, etc.)

<http://www.webis.de/research/events/pan-09>



Plagiarism detection (source retrieval / text alignment)

Authorship identification

Sexual predator identification

Quality flaw prediction in Wikipedia

+ **Univ. of Lugano** (Fabio Crestani &

Giacomo Inches)

<http://www.webis.de/research/events/pan-12>

PAN @ FIRE

- ▶ 2011 Task on Cross Language Indian Text Reuse (CL!TR)
- ▶ 2012 Task on Cross-Language Indian News Story Search (CL!NSS)
- ▶ 2013 Task on Cross-Language Indian News Story Search (CL!NSS)
- ▶ 2014 Task on Detection of SOurce COde Re-use (SOCO)
- ▶ 2015 Task on Cross-Language Detection of SOurce COde Re-use??

The word 'SOCO' is written in a large, black, handwritten-style font. Below the first 'O', there is a curved arrow that starts under the 'S' and points towards the second 'O'.

<http://www.dsic.upv.es/grupos/nle/soco>

PAN: Lab @ CLEF 2013

Registered teams: 92 (from 38 countries) with more than 150 different team members; per tasks:

Plagiarism detection: 38

Authorship identification: 47

Author profiling: 55

Virtual machines (Linux / Windows): 35

Early bird evaluation (15th March): 25; per task:

Plagiarism detection: 11

Authorship identification: 9

Author profiling: 5

Softwares submitted: 62; per task:

Plagiarism detection: 19

Authorship identification: 18

Author profiling: 21



PAN: Lab @ CLEF 2014



Tasks:

- Plagiarism detection (source retrieval / text alignment)
- Authorship identification
- Author profiling (age and gender identification in social media; in English and Spanish) **with news!!**

+ **University of Antwerp** (Walter Daelemans)

URL: <http://pan.webis.de/>

PAN: Lab @ CLEF 2014



News on Author Profiling

PAN-Replab Collaboration

Two
complementary
perspectives
of **Author
Profiling**

PAN virtual machines
for **RepLab** participants
TIRA platform @ Weimar

- All participants with the same **computing power**
- Increases participants **engagement**
- Improves **Sustainability, Replicability** and **Reproducibility**
- Allows **cross-year evaluations**

New Datasets

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

PAN: Lab @ CLEF 2014

Current statistics

Registered teams: 91 (from 37 countries)

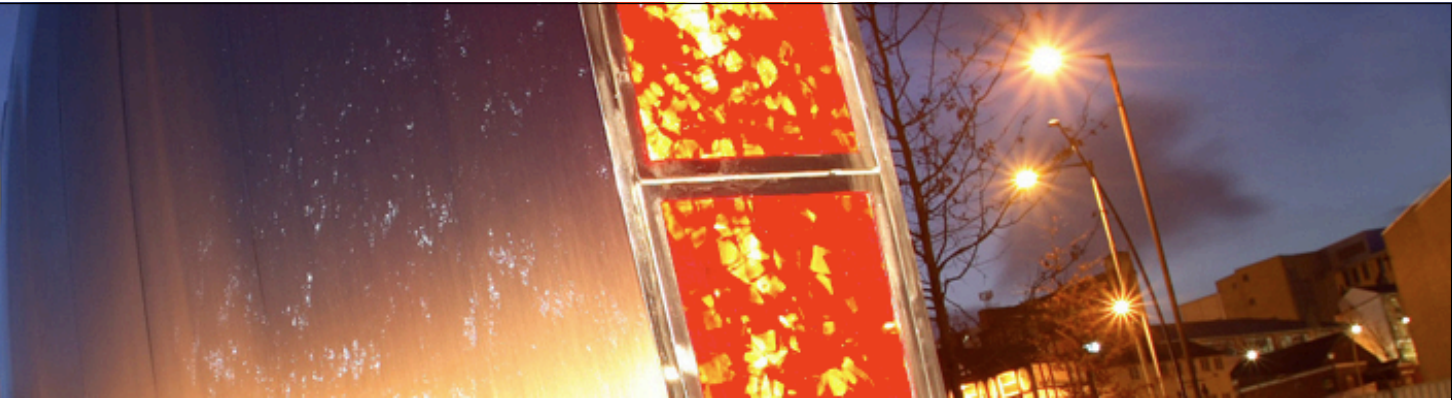
Plagiarism detection: 49

Authorship identification: 49

Author profiling: 46

Virtual machines (Linux / Windows): 40





See you there!
15-18 Sept