

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/281448461>

QUALITY OF WEB DATA : A STATISTICAL APPROACH FOR FORENSICS

Conference Paper · September 2015

READS

51

4 authors:



Vito Santarcangelo

Centro Studi , Buccino (SA), Italy

40 PUBLICATIONS 15 CITATIONS

SEE PROFILE



Antonio Buondonno

University of Naples Federico II

5 PUBLICATIONS 0 CITATIONS

SEE PROFILE



Angelo Romano

Politecnico di Torino

8 PUBLICATIONS 0 CITATIONS

SEE PROFILE



Egidio Cascini

Accademia Italiana Sei Sigma

14 PUBLICATIONS 3 CITATIONS

SEE PROFILE

QUALITY OF WEB DATA : A STATISTICAL APPROACH FOR FORENSICS

V. Santarcangelo^{1,3}, A. Buondonno³, A. Romano³, E. Cascini²

1. Centro Studi S.r.l., 2. Accademia Italiana del Sei Sigma, 3. iInformatica S.r.l.s.

vito.santarcangelo@centrostudi.biz

Abstract:

Today web data are one of the most important sources in all of the world interconnected through the internet network. Digital information, growing day by day, are becoming more and more important for people and business. At same time, the growing of fake data indubitably decreases the overall quality, that is an indispensable requisite for their use in the area of forensics applications, for example. In a previous work, namely WEB MISINFORMATION: A TEXT-MINING APPROACH FOR LEGAL ACCEPTED FACT a solution to this important topic has been proposed, in order to allow a safe use of web information in legal contexts. Scope of the present work is providing a system to evaluate the *quality of web data* from a more general view point, i.e. taking into consideration different gathering approach, such as images, video and audio. Another work important target is improving the *notoriety system* of the paper cited above, considering the state of the art of the *text similarity* approaches, choosing the best one for the system and developing a robust statistical approach for a scoring system. This approach will provide a system to define the quality of data for forensics scopes.