

# WEB MISINFORMATION: A TEXT-MINING APPROACH FOR LEGAL ACCEPTED FACT

V. Santarcangelo<sup>1,4,5</sup>, A. Buondonno<sup>1</sup>, G. Oddo<sup>1,5</sup>, F. P. Staffieri<sup>2</sup>

N. Santarcangelo<sup>3,4</sup>, M. Maragno<sup>4</sup>, F. Trentinella<sup>6</sup>

1. iInformatica S.r.l.s., 2. Studio Legale Staffieri, 3. BNG SpA

4. Ordine degli Ingegneri della Provincia di Matera, 5. Centro Studi S.r.l.

6. Università Telematica E-campus

[vito.santarcangelo@centrostudi.biz](mailto:vito.santarcangelo@centrostudi.biz)

*Choice and preference  
analysis for quality  
improvement and seminar  
on experimentation*

10/07/2015 - BARI

*Centro Studi*

Process Development & Applied Research

 gruppo orizzonti holding



# WEB DATA

More and more information appear in the web every day, in our world interconnected by the web.

This amount of information is also known as “Web Big Data”.



# MISINFORMATION AND DISINFORMATION

However, the web information can present also inaccurate information (web information spoofing).

This kind of problem is growing day by day.

- Misinformation is the unintentional inaccurate information
- Disinformation is the intentional inaccurate information

These two problems introduces lot of noise in the analysis and results of “Big Data”.

# TOOL FOR DISINFORMATION

www.fodey.com/generators/newspaper/snippet.asp  
**The Newspaper Clipping Generator**



Download your Image

**Name of the newspaper:**

ASA News

**Date:**

Tuesday, July 08, 2015

**Headline:**

<http://mediafetcher.com/>

## CELEBRITY FAKE NEWS HOAX GENERATOR



FAKEAWISH.COM - PUNKING MASS MEDIA SINCE 2002

[home](#) [disable name](#) [news](#) [contact](#) [facebook](#) [twitter](#)

**SUCCESS!**

here's your link... go share it and fool some friends!

[Andrea Pirlo breaks both legs in car accident](#)

*note: be sure to check the city... you can simply edit the name of the city in the URL to anything you want and it will work!*

[andrea.pirlo.globalassociatednews.com/news/athlete/retirement/NJ/football/](http://andrea.pirlo.globalassociatednews.com/news/athlete/retirement/NJ/football/)

### Trending Topics

Mckinney Texas WWDC 2015 Apple National Bestfriend Day Father S Day Brandon Sp Moss MLB draft Tim Curry Best Friends Day Women S World Cup

## Andrea Pirlo Announces Unexpected Retirement | PRO FOOTBALL NEWS

[Mi piace](#) [Condividi](#) [Tweet](#) 0

**THIS STORY IS STILL DEVELOPING...**

NJ July 8, 2015 30 minutes ago...

Andrea Pirlo's sudden and unexpected announcement of retirement sends social media trending wires into a frenzy with unanswered questions and speculation.



# TOOL FOR DISINFORMATION



Fill out the form below to generate a fake news page that you can use to fool your friends!

Note: This is only intended as a **practical joke** and should not be taken seriously!

## Select a Topic

Wanted For Tax Evasion ▼

## Victims's First Name

Francesco

## Victims's Last Name

Zalone

## NEW Victim's Facebook ID What's this?

CheccoZaloneOriginal (optional)

## Location

Bari, BA

FINISH

PREVIEW

SHARE IT WITH YOUR FRIENDS

SHARE

Facebook Twitter Email

<http://www.noob.co.in/francesco/zalone/bari%2c%20ba/business/ch>

## FRANCESCO ZALONE WANTED FOR TAX EVASION



Available picture of Francesco Zalone

Francesco Zalone is wanted by the IRS for owing over 75,000 in back taxes. On his web site, Zalone has not filed a federal tax return in five years.

The penalty for misdemeanor tax evasion is one year in prison for each year, with a maximum of three years imprisonment. Zalone currently faces a three-year sentence in district court. He is also being charged with more serious felony charges of tax evasion conspiracy.

"This should send a message to everyone who thinks they can get away with tax evasion because they have extremely low income," said Assistant Attorney General John J. Donohue of the Justice Department's Tax Division.

Posted at 1:22PM on 07/7

SHARE

Facebook Twitter Email

Local Board Cuts More Jobs in Bari, BA

Teachers in the Bari, BA school district are

Rainy Season is Over For Local Farmers

The farming community is breathing a sigh of



# TOOL FOR DISINFORMATION




<http://bit.ly/1DoHA3B>

Enter a site to Remix

URL: <http://analysisquality.stat.unij>



Convegno

 **CONVEGNO ANNULATO!! Misura e analisi di scelte  
e preferenze per migliorare la qualità  
e seminario sulla sperimentazione**

Palazzo del Demanio, Matera  
9-10 luglio 2015

# A LEGAL PROBLEM

Legal context about “**ACCEPTED FACT**” in Italian legislation

Tribunale di Mantova, ord. 16 maggio 2006 – G.U. Dr. Laura De Simone.

**Consulenza tecnica d’ufficio – Acquisizione di notizie e documenti non autorizzati – Nullità – Sussistenza.**

Deve essere dichiarata la nullità della consulenza tecnica d’ufficio nella parte in cui riporti notizie e documentazione in mancanza dei presupposti di cui all’art. 198 c.p.c. (consenso delle parti all’esame di documenti non prodotti in causa) o dell’art. 2711 c.c. (richiesta di acquisizione da parte dell’ufficio)

**Notizie reperibili in internet – Notizie di comune esperienza – Fatto notorio – Esclusione.**

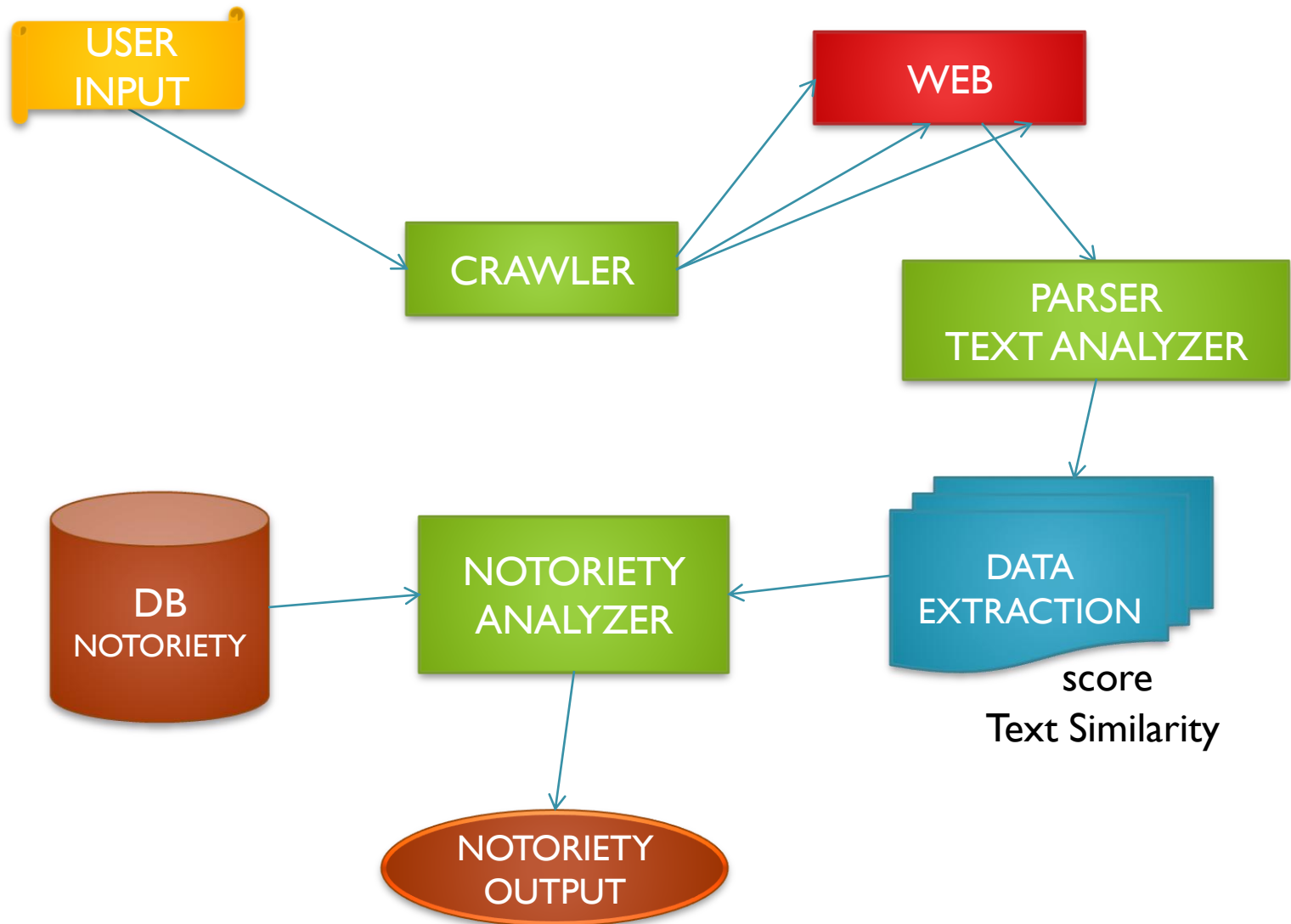
Le notizie reperibili in internet non costituiscono di per sè nozioni di comune esperienza, secondo la definizione dell’art. 115 ult. co. c.p.c., norma che, derogando al principio dispositivo, deve essere interpretata in senso restrittivo. Può, infatti, ritenersi "notorio" solo il fatto che una persona di media cultura conosce in un dato tempo e in un dato luogo, mentre le informazioni pervenute da internet, quand’anche di facile diffusione ed accesso per la generalità dei cittadini, non costituiscono dati incontestabili nelle conoscenze della collettività.

# OUR IDEA

- TO DEVELOP A TEXT MINING SYSTEM CRAWLER BASED TO INVESTIGATE THE “NOTORIETY” OF THE WEB DATA
- WEB NOTORIETY IS AN IMPORTANT TARGET TO REACH IN LEGAL MATTER



# OUR NOTORIETY SYSTEM



# NOTORIETY KNOWLEDGE BASE

- Database of over 1000 entries (shared for improvement)
- Score from +3.0 (better notoriety) to -3.0 (worst notoriety)
- .edu / .gov.it / .int / .museum (Score +3.0)

WEBSITE	NOTORIETY	APPLICATION FIELD
www.ansa.it	+3.0	News
www.wikipedia.it	+2.0	General
nonciclopedia.wikia.com	-3.0	Funny
www.istruzione.it	+3.0	Institutional
www.uniba.it	+3.0	Institutional
www.sportnotizie24.it	+2.0	News
www.barzellette.net	-3.0	Funny
www.bufale.net	-3.0	Funny
www.altervista.org	-1.0	Web Hosting

# OUR METRIC

- **INFORMATION**

$$N > +1.0$$

- **MISINFORMATION**

$$-1.0 \leq N \leq +1.0$$

- **DISINFORMATION**

$$N < -1.0$$

$$N = \frac{\sum_{i=1}^n W(x_i) \times \phi(x_i)}{n}$$

n is the number of website  
extracted

where

$W(x_i)$  DB NOTORIETY WEIGHT

$\phi(x_i)$  TEXT ANALYZER SCORE

# AN EXAMPLE

- SHOCK – Rakitic, Suarez, Neymar POSITIVI AL DOPING – Chiesta Squalifica – Coppa alla Juventus?

Crawler makes an extraction from the web, Parser and Text Analyzer recover relevant information

- <http://corrieredel mattino.altervista.org> (-1) [1]
- <http://www.bufale.net> (-3) [1] **DISINFORMATION SOURCE**
- <http://www.parestrano.it> (-3) [1] **DISINFORMATION SOURCE**
- <http://www.sportnotizie24.it> (+2) [0.2]

Notoriety score is **-1,65**

The Notoriety Output about the input news is **DISINFORMATION**

# FUTURE WORKS

- DEVELOPMENT of a **NOTORIETY Plugin for Browser** (for a real-time news analysis)
- IMPROVE the **NOTORIETY DATASET** through visitors' feedbacks (by a browser plugin)



# REFERENCES

For more information  
and dataset visit



- [http://www.researchgate.net/profile/Vito\\_Santarcangelo](http://www.researchgate.net/profile/Vito_Santarcangelo)
- [http://www.researchgate.net/publication/278685425\\_WEB\\_MISINFORMATION\\_A\\_TEXT-MINING\\_APPROACH\\_FOR\\_LEGAL\\_ACCEPTED\\_FACT](http://www.researchgate.net/publication/278685425_WEB_MISINFORMATION_A_TEXT-MINING_APPROACH_FOR_LEGAL_ACCEPTED_FACT)