

Exploiting messages posted on online social networks for saving lives



Sotiria Giannitsari

June 2016

Ψηφιακά Κοινωνικά Δίκτυα: Αναρτήσεις που σώζουν ζωές



ΟΠΑ
ΑΥΕΒ

Σωτηρία Γιαννίτσαρη

Εργαστήριο Ασφάλειας Πληροφοριών & Προστασίας
Κρίσιμων Υποδομών (INFOSEC Laboratory)
Τμήμα Πληροφορικής | Οικονομικό Πανεπιστήμιο Αθηνών



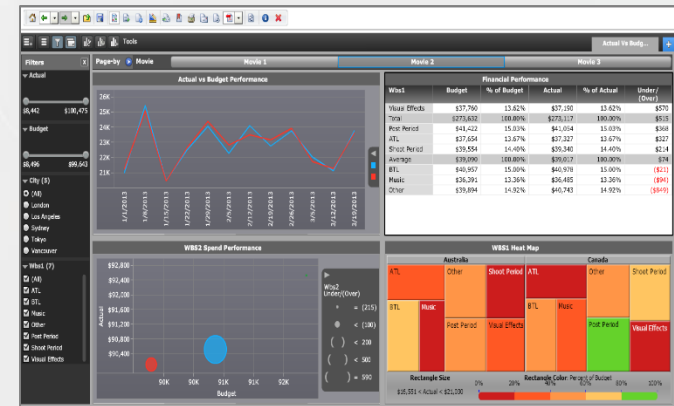
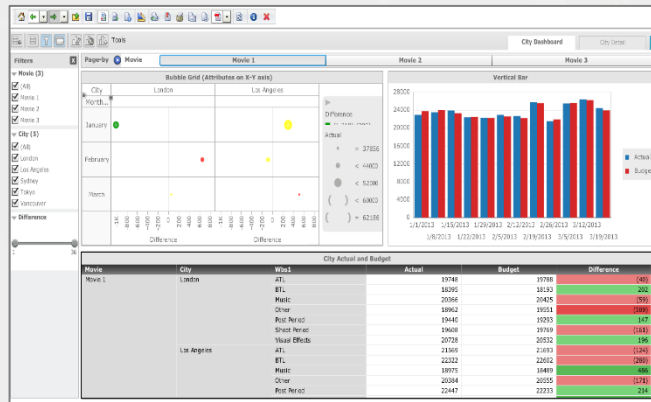
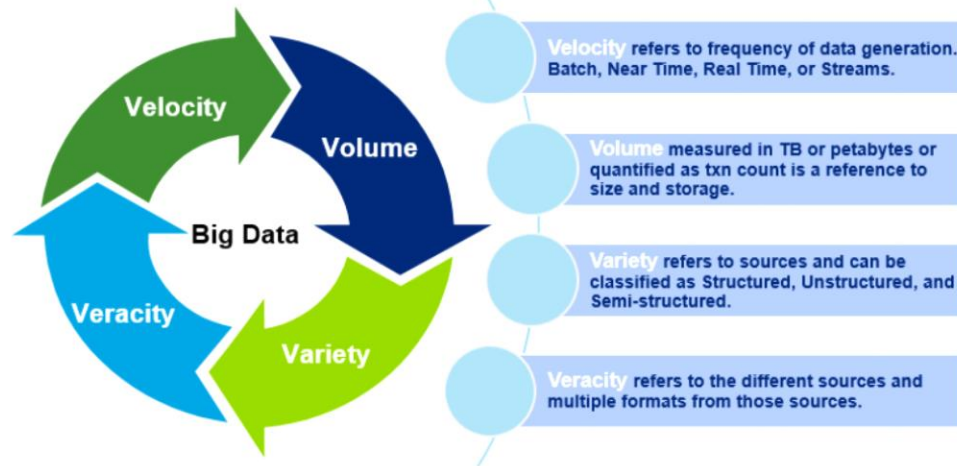
InfoSec

Big Data

Πακτώλός δεδομένων στη διάθεση του ερευνητή

4V: Velocity (Ρυθμός), Volume (Όγκος), Variety (Ποικιλία), Veracity (Εγκυρότητα)

A whole new class of inputs driven by an explosion in the volume of data being created by social media, machines, intelligent devices, and other applications that is creating new challenges and opportunities.



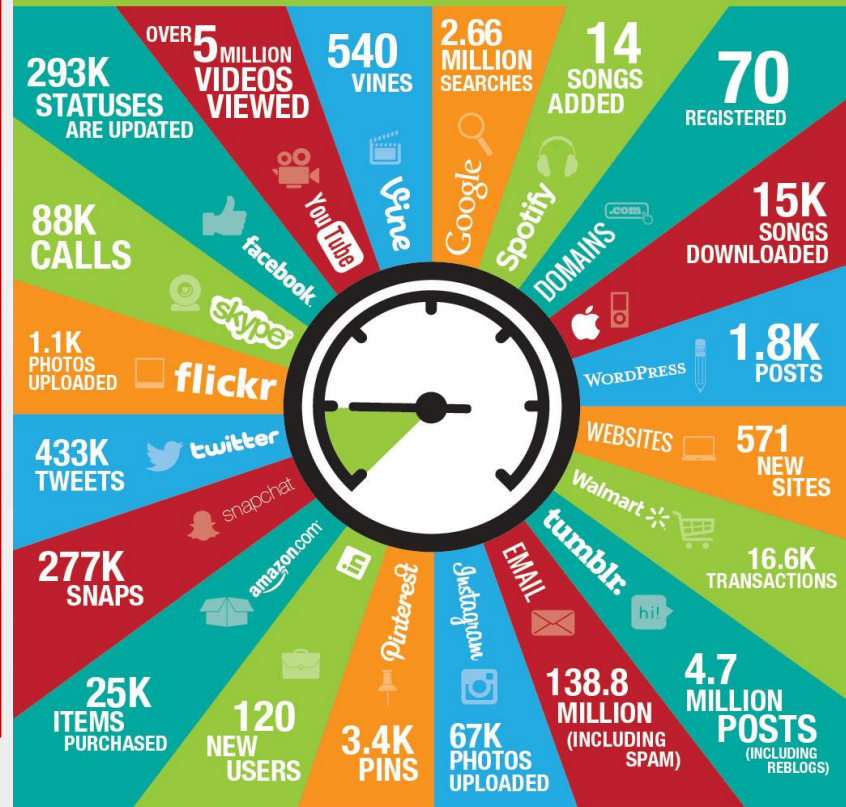
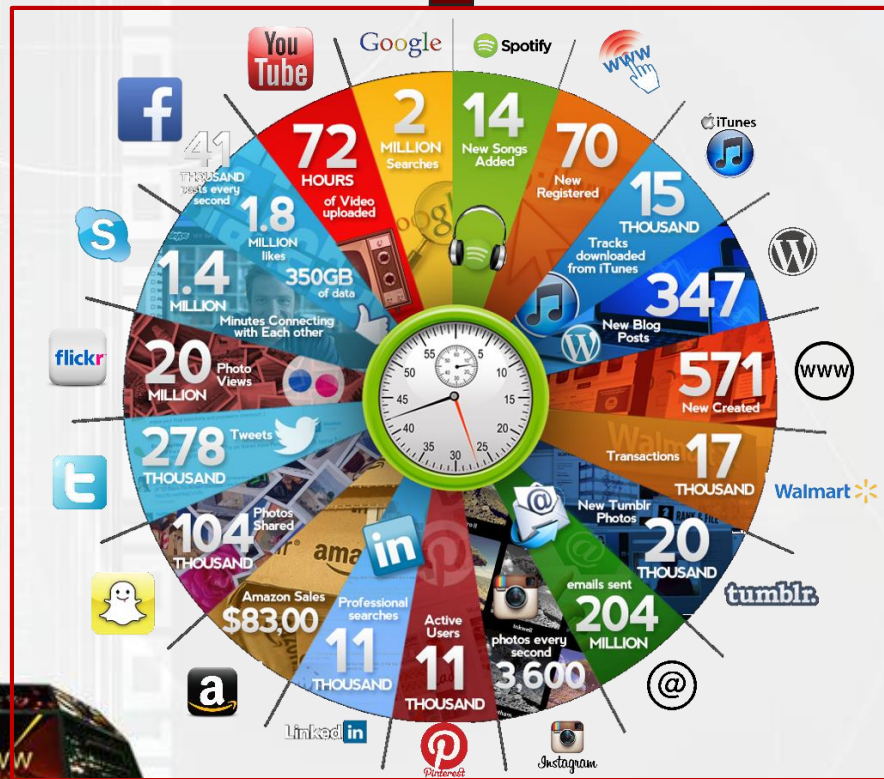
Αχανή πεδία έκφρασης και επικοινωνίας

Internet, Web 2.0, Online Social Networks

qmee.com

ONLINE IN
60
SECONDS
A YEAR LATER

WE ALL KNOW ACTIVITY ON THE INTERNET ON A DAILY BASIS MOVES AT LIGHTNING SPEED, BUT THERE'S SOMETHING ABOUT HAVING THE NUMBERS IN FRONT OF YOU THAT MAKES IT JUST A LITTLE MORE FASCINATING. LAST YEAR, WE PUBLISHED THE INFOGRAPHIC "ONLINE IN 60 SECONDS" AND WE THOUGHT IT WOULD BE GREAT TO TAKE A LOOK AT HOW THE NUMBERS EVOLVE, IN JUST ONE YEAR.



DATA
www.internetstats.com
www.thesocialstats.com
www.tumblr.com
www.gowalla.com
www.amazon.com
www.etsy.com
www.flickr.com
www.widert.com
www.mashable.com

DESIGN BY NoLimitAgency

Ερευνητικό ερώτημα και όραμα

Αποτροπή απώλειας ανθρώπινης ζωής



- ⇒ Διατυπώνονται στα Ψηφιακά Κοινωνικά Δίκτυα **προειδοποιήσεις** ή/ και **απειλές για αφαίρεση ανθρώπινης ζωής** (αυτοχειρία);
- ⇒ Είναι δυνατός ο αποτελεσματικός **εντοπισμός** των απειλών;
- ⇒ Είναι δυνατή η **συλλογή και ανάλυση** των αναρτήσεων;
- ⇒ Είναι δυνατή η **αξιολόγηση της αξιοπιστίας** των αναρτήσεων;
- ⇒ Υπάρχουν προβλήματα **νομιμότητας** αυτής της διαδικασίας;
- ⇒ Είναι δυνατόν να **αξιοποιηθεί** στην πράξη αυτή η διαδικασία;

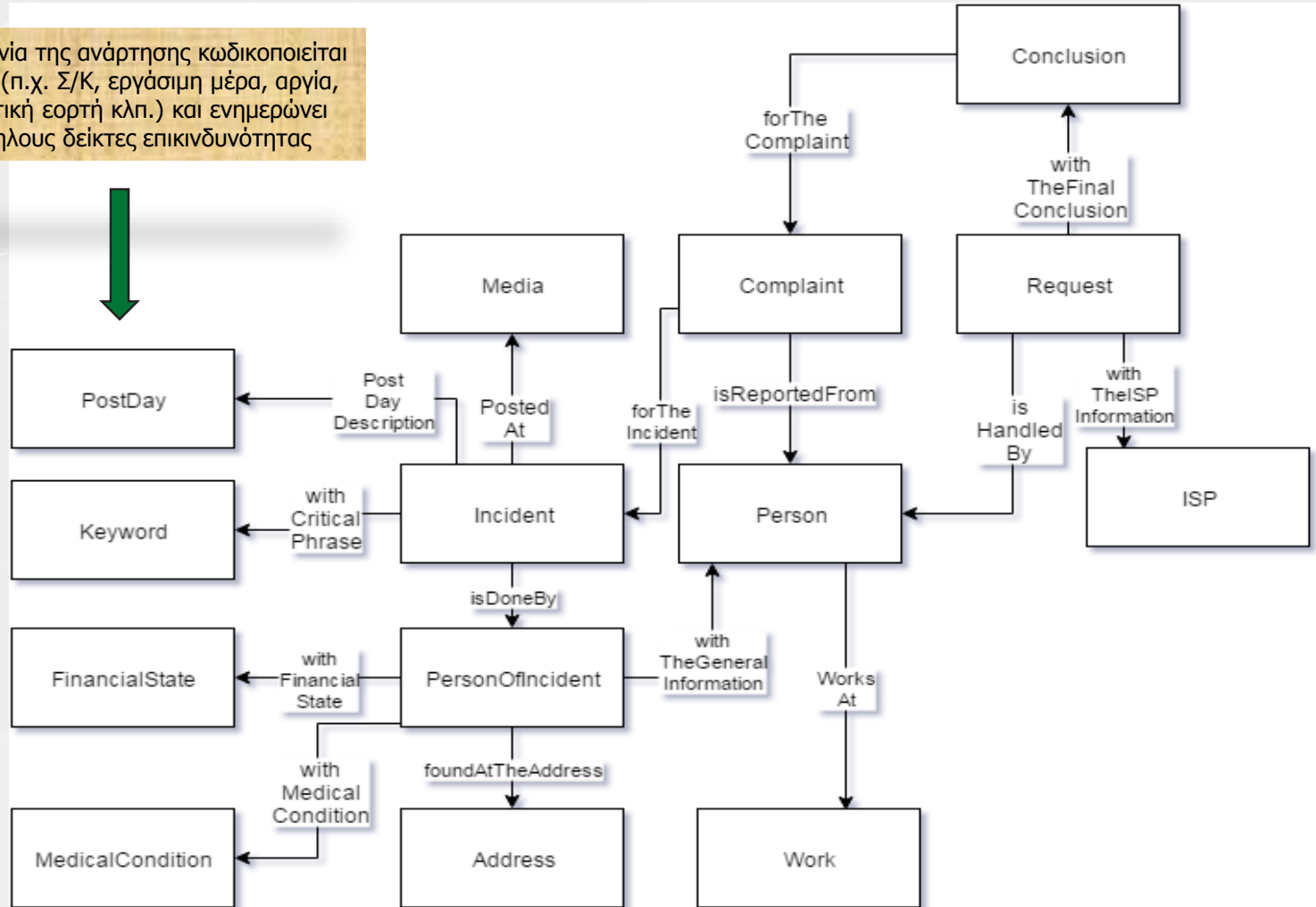




Σχεδιασμός Βάσης Δεδομένων

Αρχική αρχιτεκτονική & συσχετίσεις δεδομένων

Η ημερομηνία της ανάρτησης κωδικοποιείται αυτόματα (π.χ. Σ/Κ, εργάσιμη μέρα, αργία, θρησκευτική εορτή κλπ.) και ενημερώνει κατάλληλους δείκτες επικινδυνότητας





Διαχείριση Κρίσιμων Αναρτήσεων

Η επικρατούσα αντιμετώπιση του προβλήματος


 Help Centre

How do I report suicidal content to Facebook?

If someone you know is in immediate physical danger, please contact your local law enforcement or a suicide hotline immediately.

We also ask that you report suicidal content to us so we can reach out to this person with information that may be helpful to them.

Learn more about how you can support a friend who's posted something about suicide or self-injury on Facebook.

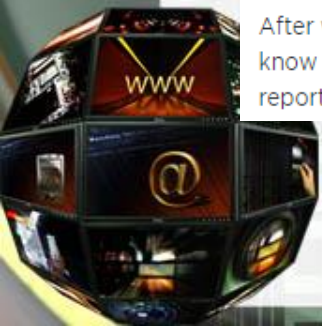
 Help Center

Dealing with self-harm and suicide

If you or someone you know is at risk of self-harm or suicide, you should seek help as soon as possible by contacting agencies specializing in crisis intervention and suicide prevention. Also [alert the team](#) devoted to handling threats of self-harm or suicide if you encounter such threats on Twitter.

Twitter's approach to self-harm and suicide threats

After we assess a report of self-harm or suicide, Twitter will contact the reported user and let him or her know that someone who cares about them identified that they might be at risk. We will provide the reported user with available online and hotline resources and encourage them to seek help.

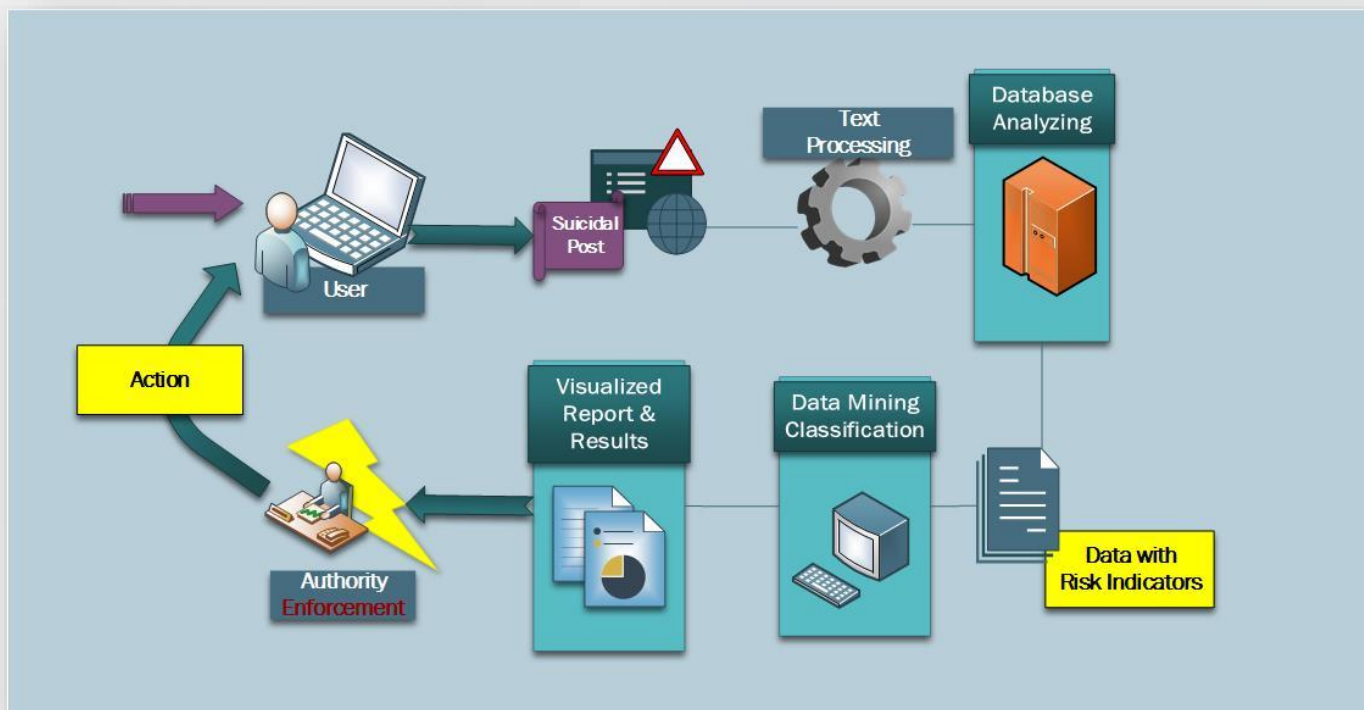




Διαχείριση Κρίσιμων Αναρτήσεων

Η νέα αντιμετώπιση για αποτροπή αυτοκτονιών

- ✓ Online real-time critical post discovery
- ✓ Automated process of data
- ✓ Efficient and faster classification
- ✓ Quick and accessible results
- ✓ Immediate action that can save human Lives





Συμβολή των ΤΠΕ στην αποτροπή αυτοκτονιών (Καταρχήν) Δυνατότητες και περιορισμοί

- ✓ Στα Ψηφιακά Κοινωνικά Δίκτυα αναρτώνται **προειδοποιήσεις και απειλές** αφαίρεσης ανθρώπινης ζωής (πχ. **πρόθεση αυτοχειρίας**).
- ✓ Ο **έγκαιρος** και **αυτόματος εντοπισμός** των απειλών είναι τεχνολογικά εφικτός (**Data Mining**).
- ✓ Η **αυτόματη συλλογή** και **ανάλυση** των αναρτήσεων είναι τεχνολογικά εφικτή (**Social Media Intelligence**).
- ? Είναι (καταρχήν) τεχνολογικά δυσχερής η **αυτόματη αποτίμηση του συναισθήματος** μιας ανάρτησης (**Sentiment Analysis**).
- ? Υπάρχουν κάποια (καταρχήν) αντιμετώπισιμα ζητήματα διασφάλισης **δικαιωμάτων** των χρηστών (**Privacy**).
- ? Η διαδικασία είναι δυνατόν (καταρχήν) να αξιοποιηθεί, σε συνεργασία με **αρμόδιες υπηρεσίες** (πχ. **Ελληνική Αστυνομία** κλπ.).



References

1. Gritzalis D., Stavrou V., Kandias M., Stergiopoulos G., "Insider Threat: Enhancing BPM through Social Media", in *Proc. of the 6th IFIP International Conference on New Technologies, Mobility and Security (NMTS-2014)*, Springer, UAE, 2014.
2. Gritzalis D., "Insider threat prevention through Open Source Intelligence based on Online Social Networks", Keynote address, *13th European Conference on Cyber Warfare and Security (ECCWS-2014)*, Greece, 2014.
3. Gritzalis D., Kandias M., Stavrou V., Mitrou L., "History of Information: The case of Privacy and Security in Social Media", in *Proc. of the History of Information Conference*, Law Library Publications, Greece, 2014.
4. Kandias M., Mitrou L., Stavrou V., Gritzalis D., "Which side are you on? A new Panopticon vs. privacy", in *Proc. of the 10th International Conference on Security and Cryptography (SECRYPT-2013)*, pp. 98-110, Iceland, 2013.
5. Kandias M., Galbogini K., Mitrou L., Gritzalis D., "Insiders trapped in the mirror reveal themselves in social media", in *Proc. of the 7th International Conference on Network and System Security (NSS-2013)*, pp. 220-235, Springer (LNCS 7873), Spain, 2013.
6. Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in *Proc. of the 6th International Conference on Critical Infrastructure Security (CRITIS-2011)*, pp. 93-103, Springer (LNCS 6983), United Kingdom, 2013.
7. Kandias M., Stavrou V., Bozovic N., Mitrou L., Gritzalis D., "Can we trust this user? Predicting insider's attitude via YouTube usage profiling", in *Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (ATC-2013)*, pp. 347-354, IEEE Press, Italy, 2013.
8. Kandias M., Stavrou V., Bosovic N., Mitrou L., Gritzalis D., "Proactive insider threat detection through social media: The YouTube case", in *Proc. of the 12th Workshop on Privacy in the Electronic Society (WPES-2013)*, pp. 261-266, ACM Press, Germany, 2013.
9. Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in *Proc. of the 6th International Workshop on Critical Infrastructure Security (CRITIS-2011)*, Bologna S., et al (Eds.), pp. 93-103, Springer (LNCS 6983), Switzerland, 2011.
10. Kandias M., Mylonas A., Virvilis N., Theoharidou M., Gritzalis D., "An Insider Threat Prediction Model", in *Proc. of the 7th International Conference on Trust, Privacy, and Security in Digital Business (TrustBus-2010)*, pp. 26-37, Springer (LNCS-6264), Spain, 2010.
11. Mitrou L., Kandias M., Stavrou V., Gritzalis D., "Social media profiling: A Panopticon or Omnipticon tool?", in *Proc. of the 6th Conference of the Surveillance Studies Network*, Spain, 2014.
12. Stavrou V., Kandias M., Karoulas G., Gritzalis D., "Business Process Modeling for Insider threat monitoring and handling", in *Proc. of the 11th International Conference on Trust, Privacy & Security in Digital Business (TRUSTBUS-2014)*, pp. 119-131, Springer (LNCS 8647), Germany, 2014.