

A witch-hunt story: Revealing citizens political beliefs via SOCMINT



Miltos Kandias
November 2015

Κυνήγι μαγισσών τον 21ο αιώνα: Εντοπισμός πολιτικών πεποιθήσεων μέσω SOCMINT



4th International e-Life Congress
Αθήνα, Νοέμβρης 2015



ΟΠΑ
ΑΥΕΒ

Μιλτιάδης Κάνδιας

Information Security & Critical Infrastructure Protection Laboratory
Dept. of Informatics | Athens University of Economics & Business

Περιεχόμενα

- ⇒ Δημόσια Διαθέσιμα Δεδομένα
- ⇒ Δυνατότητες SOCMINT
- ⇒ Αρχιτεκτονική Συστήματος
- ⇒ Εντοπισμός Πολιτικών Πεποιθήσεων
- ⇒ Παρατηρήσεις
- ⇒ Συμπεράσματα



Δημόσια διαθέσιμα δεδομένα στα Online Social Networks

- Η δομή των Online Social Networks επιτρέπει προσωποποιημένη χρήση
- Η πλειονότητα των δεδομένων στο διαδίκτυο παράγονται από χρήστες
- Οι χρήστες αποκαλύπτουν, αβίαστα, στοιχεία της προσωπικότητάς τους
- Υπάρχουν ποικίλα κίνητρα χρήσης τους (επαγγελματικά, διασκέδαση, επικοινωνία)
- Οι χρήστες αναπαράγουν τη συμπεριφορά τους (και) στο Διαδίκτυο

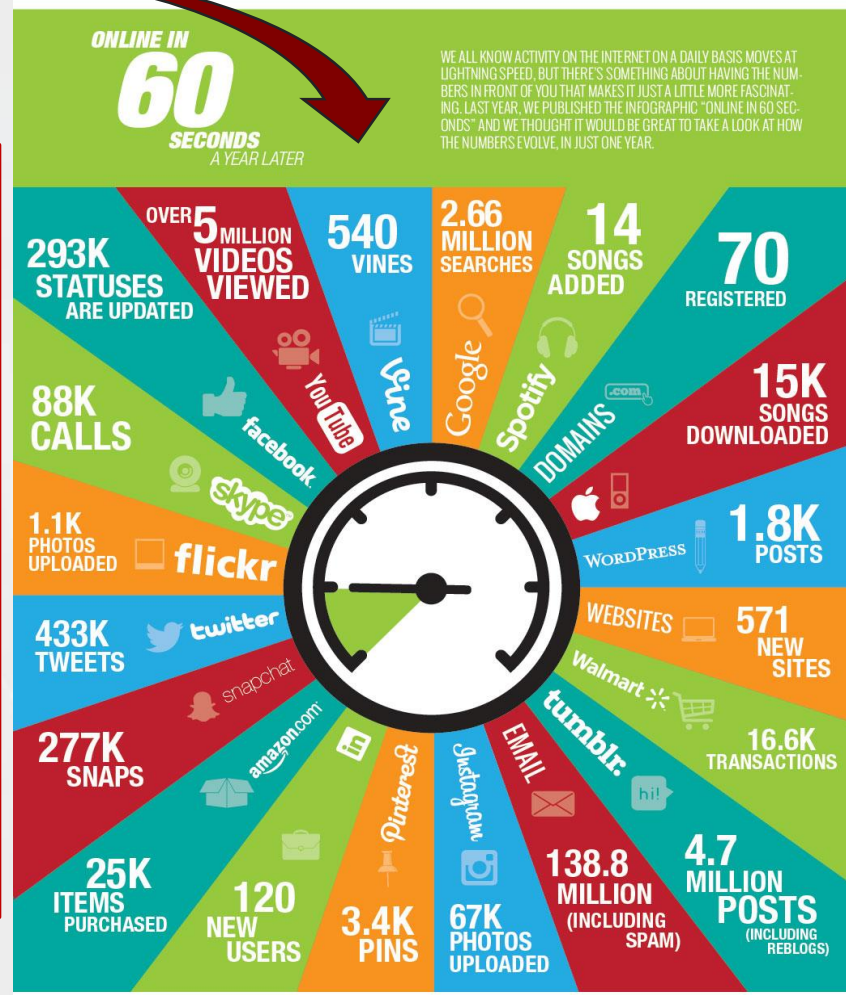
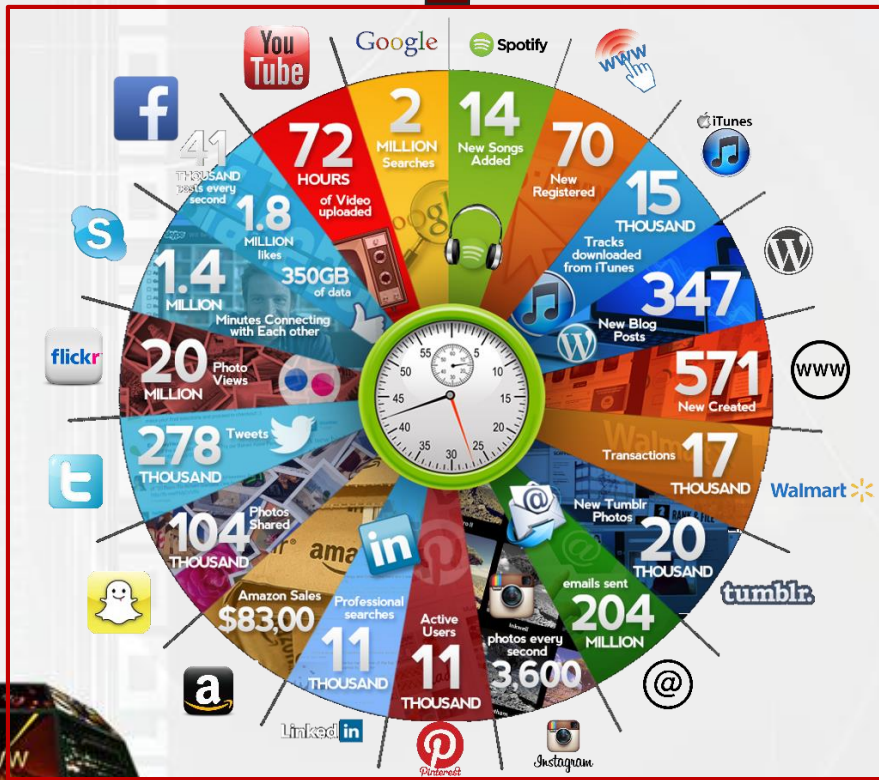
SOCMINT

Καθημερινά παράγεται ένας τεράστιος όγκος νέων δεδομένων τύπου **Social Media Intelligence (SOCMINT)**, πληροφορίες παραγόμενες από δημόσια διαθέσιμα δεδομένα, που μπορούν να αξιοποιηθούν για την αντιμετώπιση ειδικών αναγκών πληροφόρησης, μεταδιδόμενες έγκαιρα και σε κατάλληλους αποδέκτες.



Web 2.0 and Online Social Networks (OSN)

qmee.com



DATA
www.internetstats.com
www.facebook.com
www.tumblr.com
www.guru.com
www.amazon.com
www.linkedin.com
www.flickr.com
www.walmart.com
www.mashable.com

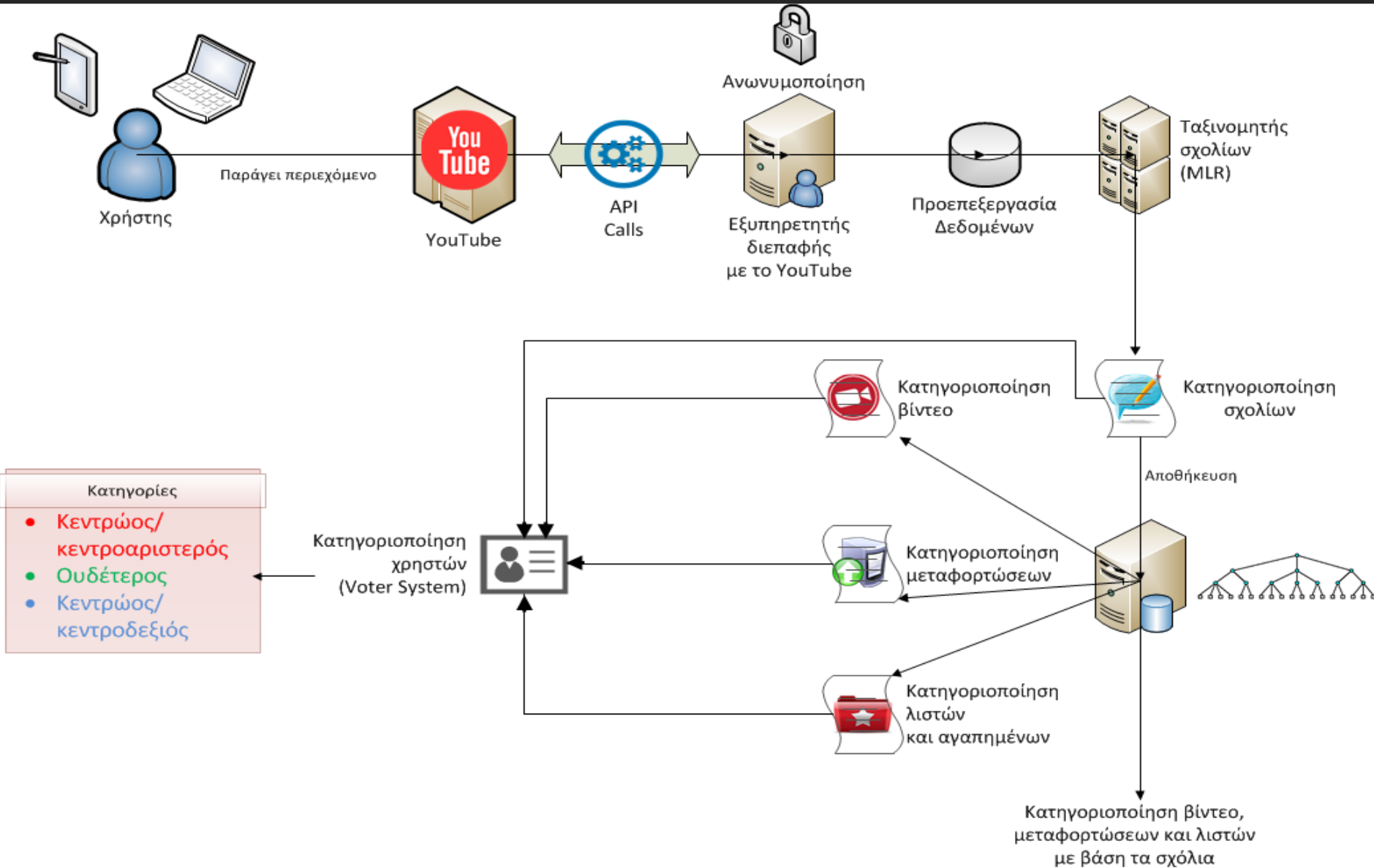
DESIGN BY NoLimitAgency

Δυνατότητες Social Media Intelligence

- Δημιουργία μορφότυπων (profiles) χρήσης και χρηστών για:
 - Προσωποποίηση περιεχομένου
 - Στοχευμένη προώθηση (marketing)
 - Βελτίωση εμπειρίας χρήσης
- Αποκάλυψη ευαίσθητων δεδομένων μέσω **άμεσων** συσχετίσεων:
 - Πολιτικές πεποιθήσεις
 - Θρησκευτικές επιλογές
 - Σεξουαλικές προτιμήσεις
- Αποκάλυψη **πολιτικών πεποιθήσεων**, μέσω των δεδομένων που αναρτώνται στο YouTube (μέσω **έμμεσων** συσχετίσεων)



Αρχιτεκτονική εντοπισμού



Δείγμα μελέτης

- **Μέσο Κοινωνικής Δικτύωσης**

YouTube

- **Δεδομένα** (Νοε 2005 - Δεκ 2012)

12.964 χρήστες, 207.377 βίντεο, 2.034.362 σχόλια

- **Πρόσφορο πεδίο εφαρμογής**

Πολιτικό περιεχόμενο, οπτικοακουστικά ερεθίσματα, συναισθηματική φόρτιση, ευρεία συμμετοχή χρηστών

- **Μέθοδοι ανάλυσης**

Γραφοθεωρητική ανάλυση (Small World Phenomenon, Indegree/Outdegree Distribution, Node Loneliness)

Ανάλυση περιεχομένου (εξαγωγή συμπερασμάτων από ανάλυση σχολίων χρηστών, λεκτική ανάλυση - Opinion Mining, Machine Learning)

Ανάλυση νέφους ετικετών (συγκέντρωση ετικετών βίντεο και απεικόνιση σε νέφος ώστε να μελετηθεί το περιεχόμενο - Tag Cloud)



Κατηγορίες μελέτης

- Ορίζονται τρεις (όλως ενδεικτικές) **κατηγορίες ταξινόμησης**:
 - «**Ριζοσπαστικοί**», «**Ουδέτεροι**», «**Συντηρητικοί**»
 - Οι παραδοχές εξαρτώνται από το κοινωνικό πλαίσιο
 - Το αντικείμενο του πειράματος είναι μια πραγματική κοινότητα χρηστών
 - Ανακλάται ένα (θεωρητικό) ιστορικοπολιτικό πλαίσιο της Ελλάδας
- Ορίζεται (όλως ενδεικτικά) η εξής **αντιστοιχίση**:
 - «**Ριζοσπαστικοί**» χρήστες:
 - Όσοι επιδεικνύουν «κεντροαριστερές» ή «αριστερές» πολιτικές απόψεις
 - «**Ουδέτεροι**» χρήστες:
 - Όσοι δεν επιδεικνύουν κάποια σαφή πολιτική άποψη
 - «**Συντηρητικοί**» χρήστες:
 - Όσοι επιδεικνύουν «κεντροδεξιές» ή «δεξιές» πολιτικές απόψεις



Μεθοδολογία

- Τα σχόλια του YouTube εντάσσονται και ταξινομούνται στις καθορισμένες κατηγορίες πολιτικών απόψεων:
 - Η ταξινόμηση των σχολίων έγινε ως ταξινόμηση κειμένου.
 - Εκπαίδευση λογισμικού με βάση το περιεχόμενο των σχολίων και την κατηγορία όπου εντάχθηκαν.
 - Ένταξη σχολίων σε κατηγορίες από ειδικούς (Κοινωνιολόγος, Πολιτικός Επιστήμονας).
- Χρήση κατάλληλου δείγματος για επιβεβαίωση εγκυρότητας μεθόδου:
 - Εισάγονται προσημασμένα δεδομένα στο σύστημα.
 - Ελέγχεται αν η αρχική σήμανση (από τον ειδικό) συμπίπτει με το συμπέρασμα.
- Η πλειονότητα των δεδομένων είναι στο ελληνικό αλφάβητο.
- Παράλληλη χρήση λατινικού αλφαβήτου από τους χρήστες (“greeklish”).
 - Αντιμετωπίστηκαν ως δύο διαφορετικές «γλώσσες» και τα δεδομένα εκπαίδευσης των δύο «γλωσσών» ενώθηκαν σε έναν ενιαίο ταξινομητή.



Ανάλυση Δεδομένων

Small World Phenomenon

- Κάθε χρήστης απέχει ≤ 6 βήματα από κάθε άλλον χρήστη μέσα στο γράφο

Indegree Distribution

- Πολλοί χρήστες με λίγες επαφές και λίγοι με πολλές. Λίγοι χρήστες είναι δημοφιλείς (οι άλλοι ασχολούνται μαζί τους)

Outdegree Distribution

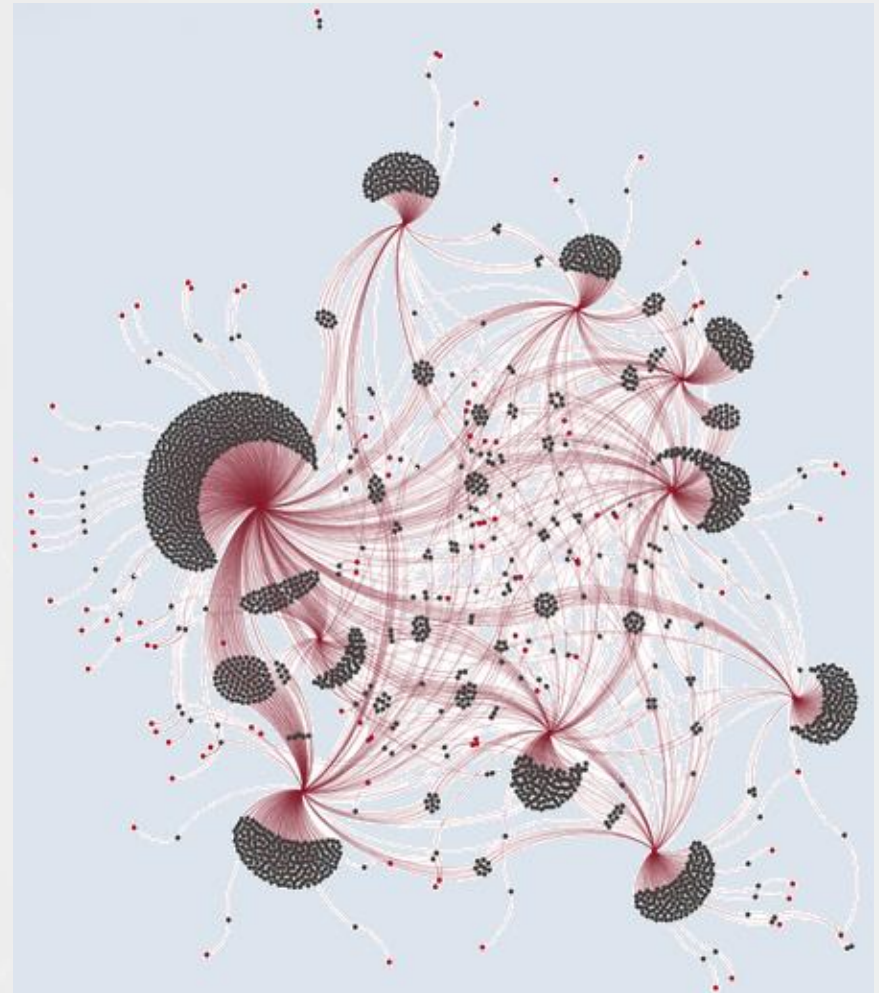
- Όλοι παράγουν περιεχόμενο. Λίγοι παράγουν το μεγάλο ποσοστό του περιεχομένου (κανόνας Pareto: Το 80% του υλικού παράγεται από το 20% των χρηστών)

Tag cloud

- Εντοπίζεται πολιτική φρασεολογία

Γράφημα χρηστών

- Οπτικοποίηση γράφου χρηστών



Αποτελεσματικότητα διαδικασίας

Κατηγορίες Χρηστών (ενδεικτικές):

«Κεντρώοι-Κεντροαριστεροί», «Ουδέτεροι»,
«Κεντρώοι-Κεντροδεξιοί»

Συνδρομή ειδικών (εντοπισμός μορφοτύπων):

Κοινωνιολόγος και Πολιτικός Επιστήμονας

Αλγόριθμος: Multinomial Logistic Regression (MLR)

Κατηγορίες	Κεντρώοι - Κεντροαριστεροί	Ουδέτεροι	Κεντρώοι - Κεντροδεξιοί
Precision	83%	91%	77%
Recall	77%	93%	78%
<u>F-Score</u>	80%	92%	77%
Accuracy		87%	

Precision: Ο αριθμός χρηστών που έχουν κατηγοριοποιηθεί ορθά, διαιρούμενος με το πλήθος των χρηστών που εντάχθηκαν σε αυτή την κατηγορία.

Recall: Ο αριθμός χρηστών που έχουν κατηγοριοποιηθεί ορθά, διαιρούμενος με το συνολικό αριθμό των χρηστών της κατηγορίας.

F-Score: Σταθμισμένος αρμονικός μέσος των Precision και Recall.

Accuracy: Ο αριθμός των ορθών κατηγοριοποιήσεων (ισούται με το πηλίκο των ορθών κατηγοριοποιήσεων δια του συνόλου των δεδομένων).

Παρατηρήσεις

2% των **σχολίων** αφορούν πολιτικά ζητήματα (0.7% Ριζοσπαστικά, 1.3% Συντηρητικά)

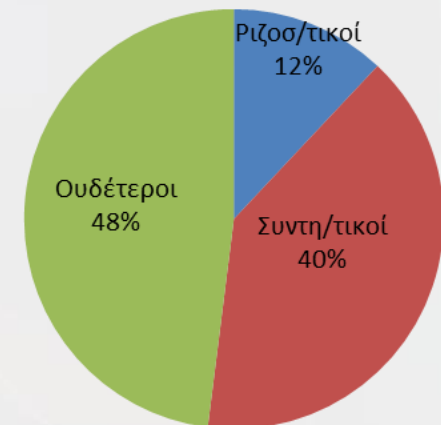
...δηλαδή, σχεδόν **41.000 σχόλια** (από τα 2.000.000) έχουν ρητό πολιτικό περιεχόμενο

7% των **βίντεο** κατηγοριοποιήθηκαν ως πολιτικά (2% Ριζοσπαστικά, 5% Συντηρητικά)

...δηλαδή, σχεδόν **14.000 βίντεο** (από τα 200.000) έχουν ρητό πολιτικό περιεχόμενο

12% των **χρηστών** εκφράζουν **Ριζοσπαστική** πολιτική τοποθέτηση και 40% **Συντηρητική**

...δηλαδή **6.760 χρήστες** «αποκαλύπτουν» την πολιτική τους τοποθέτηση



Κοινωνικοπολιτική παράμετρος

Ολοκληρωτικά Καθεστώτα



Η άνευ ελέγχου αποκάλυψη των πολιτικών πεποιθήσεων δύναται να χρησιμοποιηθεί από το Καθεστώς εναντίον αντιστασιακών και κινημάτων για τα ανθρώπινα δικαιώματα.



Οποιαδήποτε κινηματική δράση υπέρ των ανθρωπίνων δικαιωμάτων θα πρέπει να αντιταχθεί στη χρήση τέτοιων μηχανισμών από τέτοια Καθεστώτα.

Δημοκρατικά Καθεστώτα



Υπό συνθήκες περιορισμένη χρήση για εντοπισμό όσων ενεργά εχθρεύονται το Δημοκρατικό Καθεστώς.



Ακόμα και το Δημοκρατικό Καθεστώς μπορεί να επιδείξει αντίσταση σε κοινωνικές αλλαγές που πηγάζουν από κοινωνικές δράσεις διεύρυνσης των ανθρωπίνων δικαιωμάτων.



Οποιαδήποτε χρήση τέτοιων τεχνολογιών μπορεί να δικαιολογηθεί μόνο με ευρεία κοινωνική συναίνεση και έλεγχο.



Συμπεράσματα

- Η αυτόματη κατηγοριοποίηση χρηστών με βάση τις πολιτικές τους απόψεις είναι εφικτή.
- Η αποκάλυψη διάφορων τύπων ευαίσθητων δεδομένων, για πλήθος πολιτών είναι εφικτή.
- Η αναγκαία τεχνογνωσία είναι κατ' αρχήν ευρέως διαθέσιμη.
- Η αναγκαία υπολογιστική ισχύς είναι ουσιαστικά αμελητέα.
- Στατιστικώς αξιόπιστα αποτελέσματα (αλλά και false positives/negatives).
- Ενδεχόμενη εφαρμογή πρέπει να είναι απολύτως ελεγχόμενη κι εντός αυστηρών δημοκρατικών πλαισίων.



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 (NMITS-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, Athens, 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, June 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, September 2014.