

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Χόμπης Γεώργιος
Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Δ. Πλεξουσάκης**

Δευτέρα, 16/04/2018, 18:00

Αίθουσα Β108, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

**“ CAPPA: Μία Συλλογική Πλατφόρμα Ευαισθητοποίησης για την Επισημείωση
Πολιτικών Απορρήτου ”**

ΠΕΡΙΛΗΨΗ

Η μαζική εξάπλωση των ψηφιακών προϊόντων και των αντίστοιχων παραγόμενων και συλλεγμένων δεδομένων χρηστών έχουν αυξήσει τη σημασία της προστασίας της ιδιωτικότητας. Επί του παρόντος, οι επιχειρήσεις και οι οργανισμοί σε διάφορα μέρη του κόσμου υποχρεούνται από το νόμο να παρέχουν πληροφορίες σχετικά με τον τρόπο επεξεργασίας των δεδομένων των πελατών τους, συνήθως με τη μορφή εγγράφων πολιτικής απορρήτου (π.χ. ο κανονισμός γενικής προστασίας δεδομένων της ΕΕ - GDPR). Παρά το γεγονός ότι οι κανονισμοί αυτοί προσπαθούν να δώσουν τον έλεγχο των προσωπικών δεδομένων πίσω στους πολίτες, συνήθως οι χρήστες δεν εμπλέκονται σε αυτή τη διαδικασία. Δεδομένου ότι τα έγγραφα αυτά είναι συνήθως μακρά και δύσκολο

να διαβαστούν, οι χρήστες δεν είναι διατεθειμένοι να αφιερώσουν πολύ χρόνο για να τα διαβάσουν και να τα κατανοήσουν.

Μια τρέχουσα κατεύθυνση επίλυσης του προβλήματος αυτού είναι ο εμπλουτισμός των εγγράφων πολιτικής απορρήτου με επισημειώσεις είτε μέσω εμπειρογνομόνων είτε μέσω αλγορίθμων μηχανικής μάθησης. Σε αυτή τη εργασία, σχεδιάσαμε και υλοποιήσαμε μια διαδικτυακή συλλογική πλατφόρμα που επιτρέπει στους χρήστες να εξερευνούν, να επισημαίνουν και να εξετάζουν τις πολιτικές απορρήτου οποιουδήποτε ψηφιακού προϊόντος (π.χ. κινητές εφαρμογές, ιστότοποι, έξυπνες συσκευές κ.λπ.) με φιλικό τρόπο. Αυτή η πλατφόρμα αποτελεί μέρος των εργαλείων που έχουν σχεδιαστεί για την κοινότητα CAPrice, μια συλλογική πλατφόρμα ευαισθητοποίησης για την προστασία της ιδιωτικότητας.

Οι πολιτικές απορρήτου επισημαίνονται χρησιμοποιώντας ένα προκαθορισμένο σύνολο ετικετών, σχεδιασμένο για να αντικατοπτρίζουν τις ανησυχίες των χρηστών σχετικά με το τι προσωπικά δεδομένα συλλέγονται και επεξεργάζονται, από ποιους, για πόσο καιρό διατηρούνται, πώς προστατεύονται και άλλες ανησυχίες γύρω από την ιδιωτικότητα. Οι χρήστες μπορούν να συνεισφέρουν προσθέτοντας οντότητες όπως ψηφιακά προϊόντα, έγγραφα πολιτικής απορρήτου και επισημάνσεις στα έγγραφα ή αξιολογώντας οντότητες που προστέθηκαν από άλλους χρήστες. Η πλατφόρμα βοηθά και εμπλέκει τους χρήστες προς αυτή την αναζήτηση μέσω διαφόρων εργαλείων εμπλοκής (π.χ. βαθμολογίες χρηστών) και εργαλεία ανάλυσης εγγράφων (π.χ. βαθμός αναγνωσιμότητας εγγράφων πολιτικής απορρήτου). Μια επισημείωση μπορεί να θεωρηθεί έγκυρη ή λανθασμένη βάσει των ψήφων των χρηστών και του συγκεντρωτικού τους σκορ που αποκτήθηκε χρησιμοποιώντας το Wilson score interval. Ο στόχος είναι να δημιουργηθεί μια συνεργατική πλατφόρμα crowdsourcing που θα αποτελέσει σημείο αναφοράς για επισημειωμένα έγγραφα πολιτικής απορρήτου από χρήστες, για χρήστες, προγραμματιστές, ερευνητές και δημιουργούς πολιτικών απορρήτου. Προς αυτή την κατεύθυνση έχουμε σχεδιάσει ένα ReST API που παρέχει πρόσβαση στη βάση δεδομένων των ψηφιακών προϊόντων και τις επισημειωμένες πολιτικές απορρήτου τους, επιτρέποντας την εκμετάλλευση αυτών των πληροφοριών για την ανάπτυξη τρίτων εργαλείων και αλγορίθμων.

Πραγματοποιήσαμε μια βασισμένη σε χρήστες αξιολόγηση της πλατφόρμας μας, όπου είχαμε δύο ομάδες χρηστών. Σε κάθε ομάδα ζητήθηκε η επισημείωση ενός συγκεκριμένου πλήθους πολιτικών απορρήτου από το σύνολο δεδομένων OPP-115, οι οποίες είναι επισημειωμένες από έμπειρους χρήστες. Στη συνέχεια από κάθε ομάδα ζητήθηκε η αξιολόγηση των επισημειώσεων της άλλης ομάδας χρηστών και η συμπλήρωση ενός ερωτηματολογίου. Η ανάλυση των αποτελεσμάτων αναδεικνύει τη φιλικότητα της πλατφόρμας μας προς το χρήστη και ότι οι επισημειώσεις σε πολιτικές απορρήτου που συλλέχθηκαν είναι υψηλής σημασίας και ποιότητας, συγκρίσιμες με εκείνες που δημιουργήθηκαν από έμπειρους χρήστες.

Chompis Georgios
M.Sc. Thesis

Computer Science Department
University of Crete
Master's Thesis Supervisor: Professor, D. Pleksousakis

Monday, 16/04/2018, 18:00
Room B108, Computer Science Dept., University of Crete

“CAPPA: A Collective Awareness Platform for Annotating Privacy Policies”

ABSTRACT

The huge expansion of digital products and the corresponding user generated and gathered data have raised the importance of users privacy and privacy concerns. Currently, businesses and organizations around the world are enforced by law (e.g. the EU General Data Protection Regulation - GDPR) to provide information about how customers' data is being treated, usually in the form of privacy policy documents. Despite the fact that such regulations try primarily to give control back to citizens over their personal data, it's a common case that users are not engaged in this process. Since such documents are usually long and hard to read, users are not willing to spend a lot of time to read and understand them.

A current direction for addressing this problem is towards enriching privacy policy documents with annotations either through expert users or machine learning algorithms. In this thesis, we designed and implemented an online crowdsourcing platform that allows users to explore, annotate and review privacy policies of any kind of digital product (e.g. mobile applications, websites, appliances, etc.) in a friendly way. The platform is part of the tools designed for the CAPrice community, a collective awareness platform for privacy concerns and expectations.

Privacy policies are being annotated using a predefined set of tags, designed to address user concerns about what data are being collected and processed, by whom, for how long they are retained, how they are secured, and other privacy concerns. Users can contribute by adding entities like digital products, privacy policies, and annotations to documents or by reviewing entities added by other users. The platform helps and engages users towards this quest through various engagement tools (e.g. user scores) and document analysis tools (e.g. readability scores of privacy policies). An annotation can be considered as a

valid or invalid one, based on the votes of the users and their aggregated score obtained using the Wilson score interval. The aim is to provide a collaborative crowdsourcing platform that will be considered the reference for user annotated privacy policy documents, for users, developers, researchers and policy makers. Towards this direction we have designed a ReST API that provides access to the database of digital products and their annotated privacy policies. As a result, this information can be exploited for the development of third party tools and algorithms.

We conducted a user-based evaluation of our platform, where users were split in two groups. Each group was asked to annotate a specific set of privacy policies obtained from the OPP-115 dataset, which is an expert-based annotated collection of privacy policies. Then each group had to review/vote the annotations of the other group and fill in the corresponding questionnaire. The analysis of the results shows the user friendliness of our platform and that the gathered crowd-sourced privacy policy annotations are of high importance and quality, comparable to annotations created by expert users.