

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

**Μεταπτυχιακό Πρόγραμμα Σπουδών
Πληροφοριακών και Επικοινωνιακών
Συστημάτων**



Μεταπτυχιακή Διατριβή

**Τεχνικές Ανωνυμοποίησης Προσωπικών
Δεδομένων**

Φωτόπουλος Φώτιος

**Επιβλέπων Καθηγητής
Κωνσταντίνος Λιμνιώτης**

Δεκέμβριος 2016

**Μεταπτυχιακό Πρόγραμμα Σπουδών
Πληροφοριακών και Επικοινωνιακών
Συστημάτων**

Μεταπτυχιακή Διατριβή

**Τεχνικές Ανωνυμοποίησης Προσωπικών
Δεδομένων**

Φωτόπουλος Φώτιος

**Επιβλέπων Καθηγητής
Κωνσταντίνος Λιμνιώτης**

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων για απόκτηση μεταπτυχιακού τίτλου σπουδών στον Φωτόπουλο Φώτιο από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών του Ανοικτού Πανεπιστημίου Κύπρου.

Δεκέμβριος 2016

ΛΕΥΚΗ ΣΕΛΙΔΑ

Περίληψη

Η συλλογή δεδομένων μεγάλου πλήθους για την αξιοποίησή τους και για ποικίλους επιστημονικούς σκοπούς αποτελεί τα τελευταία χρόνια ένα πολύτιμο εργαλείο για την εξόρυξη γνώσης. Ωστόσο, από τη δημοσιοποίηση ή απλά διαβίβαση μεγάλου όγκου δεδομένων τίθενται θέματα παραβίασης των προσωπικών δεδομένων των ατόμων που τα δεδομένα αφορούν. Υπάρχει συχνά η λανθασμένη εντύπωση ότι απαλείφοντας τα στοιχεία ταυτοποίησης των ατόμων από τις λίστες αυτές καθίστανται ανώνυμες – κάτι η πράξη έχει καταδείξει ότι δεν ισχύει όπου σημαντικές πληροφορίες για κάποια άτομα έχουν σε διάφορες περιπτώσεις διαρρεύσει λόγω μη επαρκούς ανωνυμοποίησης. Ακριβώς για αυτούς τους λόγους έχουν προταθεί διάφορες τεχνικές για την ανωνυμοποίηση λιστών με προσωπικά δεδομένα – κάθε μια από τις οποίες με τα δικά της πλεονεκτήματα και μειονεκτήματα.

Στην παρούσα διατριβή μελετώνται και αναλύονται οι γνωστές τεχνικές ανωνυμοποίησης: γίνεται εκτενής περιγραφή του είδους των επιθέσεων που μπορεί να πραγματοποιήσει κανείς προκειμένου να άρει μια ανωνυμοποίηση, καθώς επίσης και με ποιον τρόπο η κάθε τεχνική αντιμετωπίζει ή όχι την κάθε επίθεση. Παράλληλα, διερευνάται και η απώλεια χρήσιμης πληροφορίας που προκύπτει από κάθε τεχνική, αφού κατά κανόνα επιθυμούμε τη βέλτιστη ανωνυμοποίηση αλλά με την ελάχιστη δυνατή απώλεια πληροφορίας. Προς περαιτέρω διερεύνηση των περιορισμών που υπάρχουν στις τεχνικές ανωνυμοποίησης, η έρευνά μας εστιάζει και σε ένα ρεαλιστικό παράδειγμα, εφαρμόζοντας τις τεχνικές αυτές σε μία σύγχρονη δημόσια προσβάσιμη λίστα προσωπικών δεδομένων. Καταδεικνύουμε ότι κρίσιμο ρόλο για την αποτελεσματικότητα της ανωνυμοποίησης έχει η κατανομή των τιμών που λαμβάνει το ευαίσθητο πεδίο – δηλαδή το πεδίο το οποίο χρήζει προστασίας. Η πρακτική αυτή εφαρμογή καταδεικνύει ότι όσο πιο ανομοιογενής είναι η κατανομή των τιμών του ευαίσθητου πεδίου, τόσο πιο δύσκολη είναι η επίτευξη επαρκούς ανωνυμοποίησης.

Λέξεις-κλειδιά: Ανωνυμοποίηση, προσωπικά δεδομένα, k-ανωνυμία, L-διαφορετικότητα

Summary

The collection and further processing of huge volume of data for several scientific purposes such as statistical analysis (i.e. to make decisions based on appropriate audience measurements etc.) is of high importance over the last years. However, publication or transmission of such data may give rise to several privacy concerns. Many data controllers assume that removing any personal data identifier from a list with personal data results in anonymous data; this assumption though is not correct, since several data breaches have occurred due to inappropriate anonymization procedure. Therefore, several specific anonymization techniques have been proposed to alleviate these issues – each of them having its own advantages and disadvantages.

In this thesis, all main anonymization techniques are studied and further analyzed; each of them suffices to thwart specific types of attacks, although none of them is adequate to prevent any possible attack. Therefore, the restrictions of these techniques are discussed, with regard to their resistance against several types of attacks. Moreover, apart from the anonymization, the corresponding information loss is also studied, since there is a known trade-off between the anonymity provided and the loss of useful information.

Towards evaluating the performance of these techniques, we proceed by studying a practical example; more precisely, having as starting point a known public list of personal data, we apply known anonymization techniques to estimate whether adequate anonymization can be achieved or not. Different scenarios are being considered, with respect to the distribution of the values of the sensitive attribute in this list. Our study exhibits that the anonymization that can be achieved is strongly contingent on this distribution – that is a uniform distribution corresponds to higher levels of anonymization, compared to a non-uniform distribution.

Key-words: Anonymization, personal data, k-anonymity, l-diversity

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής μου εργασίας νιώθω την ανάγκη να ευχαριστήσω θερμά τον επιβλέποντα Καθηγητή μου Δρ. Κωνσταντίνο Λιμνιώτη για την αμέριστη στήριξη και υπομονή του καθ' όλη την διάρκεια της εκπόνησής της καθώς και για την προσπάθεια του να μου μεταλαμπαδεύσει τις γνώσεις του.

Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου για την στήριξή τους που μου παρείχαν όλο αυτό το διάστημα.

Περιεχόμενα

Ανοικτό Πανεπιστήμιο Κύπρου.....	i
Μεταπτυχιακή Διατριβή.....	i
Τεχνικές Ανωνυμοποίησης Προσωπικών Δεδομένων.....	i
Περίληψη.....	v
Summary.....	vi
Ευχαριστίες.....	vii
Περιεχόμενα.....	viii
Κεφάλαιο 1.....	1
1.1 Εισαγωγή.....	1
1.2 Ιδιωτικότητα της πληροφορίας.....	1
1.3 Η έννοια των προσωπικών δεδομένων.....	3
1.4 Δομή μεταπτυχιακής διατριβής.....	4
Κεφάλαιο 2.....	5
2.1 Εισαγωγή.....	5
2.2 Περιστατικά Επιθέσεων και κακής ανωνυμοποίησης.....	6
2.2.1 Το περιστατικό της NetFlix.....	6
2.2.2 Το περιστατικό της AOL.....	7
2.2.4 Genomic Data Re-identification.....	8
2.2.5 Η περίπτωση του κυβερνήτη της Μασαχουσέτης William Weld... 	10
2.2.6 Το πείραμα του χρήστη Χ.....	11
2.4 Πόσο δύσκολη είναι η ανωνυμοποίηση.....	12
2.5 Απώλεια Πληροφορίας.....	12
2.6 Γενίκευση (Generalization).....	13
2.7 Κατάργηση (Suppression).....	14
Κεφάλαιο 3.....	17
3.1 Εισαγωγή.....	17
3.2 Κ-Ανωνυμία (K-Anonymity).....	18
3.2.3 Μειονεκτήματα Κ-Ανωνυμίας.....	20
3.2.4 Επιθέσεις στην Κ-Ανωνυμία.....	21
3.3 Τεχνική L - Διαφορετικότητας(L -Diversity).....	22
3.3.1 Οι τεχνικές Entropy L-Diversity και Recursive (c, l) Diversity.....	24
.....	24

3.3.2 Μειονεκτήματα στην L - Διαφορετικότητα	24
3.3.3 Επιθέσεις της L - Διαφορετικότητας	25
3.4 T-Εγγύτητα (T-Closeness)	27
3.5 Ανατομία (Anatomy)	27
Κεφάλαιο 4	31
4.1 Εισαγωγή	31
4.2 Εφαρμογή Τεχνικής K-Ανωνυμίας	32
4.2.1 Διαδικασία Εφαρμογής τεχνικής K-Ανωνυμίας	33
4.2.2 Η διαδικασία της Γενίκευσης	34
4.2.3 Η διαδικασία της Κατάργησης	35
4.2.4 Προσθήκη Θορύβου	35
4.2.5 Το ευαίσθητο χαρακτηριστικό	36
4.3 Εφαρμογή L - Διαφορετικότητας	36
4.3.1 Μετρήσεις στο Πεδίο 1 (Ομοιογενής Κατανομή)	37
4.3.2 Μετρήσεις στο Πεδίο 2 (Ανομοιογενής Κατανομή)	38
Κεφάλαιο 5	40
Επίλογος	40
Παράρτημα Α	41
Παράρτημα Β	42
Βιβλιογραφία	59

Κεφάλαιο 1

Εισαγωγή

1.1 Εισαγωγή

Η λέξη ανωνυμοποίηση προέρχεται από την λέξη ανωνυμία που ουσιαστικά προσδιορίζει κάποιον που δεν έχει όνομα. Η ανωνυμοποίηση είναι μία μέθοδος η οποία επεξεργάζεται ευαίσθητα προσωπικά δεδομένα με τέτοιο τρόπο έτσι ώστε να μην θίγεται η ιδιωτικότητα του ατόμου (δηλαδή να μετατρέπονται τα δεδομένα σε μορφή τέτοια ώστε να μην είναι δυνατή η αναγνώριση των προσώπων στα οποία τα δεδομένα αναφέρονται), παράλληλα όμως, στην πράξη, μας ενδιαφέρει τα δεδομένα αυτά να είναι σε τέτοια κατάλληλη μορφή που θα εξυπηρετήσουν στατιστικούς σκοπούς για το δημόσιο καλό ή για γενικότερους επιστημονικούς/ερευνητικούς σκοπούς (Alexin, 2014).

1.2 Ιδιωτικότητα της πληροφορίας

Η ιδιωτικότητα (Privacy) σαν έννοια είναι πολυδιάστατη και έχει πάρα πολλές πτυχές νομικές και μη. Σαν γενικός ορισμός, η ιδιωτικότητα μπορεί να προσδιοριστεί σαν το αναφαίρετο δικαίωμα του ατόμου στο να είναι αυτόνομο και να προστατεύει τον εαυτό του από οποιαδήποτε επίθεση που παραβιάζει το δικαίωμά του αυτό.

Οι διάφορες πτυχές της ιδιωτικότητας που έχουν προκύψει είναι μεταξύ άλλων το δικαίωμα του ατόμου να μένει μόνο, της περιορισμένης πρόσβασης που αφορά την δυνατότητα του ατόμου να συμμετέχει σε μία κοινωνία όπου δεν θα υπάρχουν άλλα άτομα ή οργανισμοί που θα συλλέγουν πληροφορίες για το άτομο αυτό, το δικαίωμα του ατόμου να αποκρύπτει πληροφορίες για τον εαυτό του, το δικαίωμα για τον έλεγχο των

πληροφοριών, για τον οποίο το άτομο έχει την πλήρη δικαιοδοσία να επιλέγει και να βάζει όρια για την ποσότητα της πληροφορίας που θα δημοσιευθεί προς τρίτους και θα αφορά αυτόν και τέλος η ιδιωτικότητα της πληροφορίας για την οποία θα γίνει πιο εκτενής αναφορά στην συνέχεια (Wikipedia-<https://en.wikipedia.org/wiki/Privacy>).

Όπως έχει ήδη αναφερθεί ο όρος ιδιωτικότητα είναι πολυδιάστατος και πολυεπίπεδος γι' αυτό και σε διαφορετικές χώρες και ηπείρους εφαρμόζεται άλλοτε με πιο αυστηρά κριτήρια και άλλοτε με πιο ήπια. Αυτά τα κριτήρια εξαρτώνται βέβαια από το σύνταγμα και τους νόμους που ισχύει στην εκάστοτε περιοχή.

Η παρούσα όμως μεταπτυχιακή διατριβή ασχολείται ειδικότερα με την ιδιωτικότητα της πληροφορίας (Information Privacy). Η ιδιωτικότητα της πληροφορίας πραγματεύεται την προστασία ευαίσθητων – υπό την έννοια των κρίσιμων/σημαντικών - πληροφοριών οι οποίες βρίσκονται σε μεγάλες βάσεις δεδομένων. Η ιδιωτικότητα της πληροφορίας μπορεί να εφαρμοστεί με αρκετούς τρόπους. Είτε μέσω της κρυπτογράφησης (Encryption), της αυθεντικοποίησης (Authentication), είτε και με την κάλυψη των δεδομένων εξασφαλίζοντας έτσι ότι τα δεδομένα δεν θα είναι διαθέσιμα σε μη εξουσιοδοτημένα άτομα.

Η ιδιωτικότητα της πληροφορίας χωρίζεται σε τρεις κατηγορίες ανάλογα τον τύπο των δεδομένων που προστατεύει και είναι οι εξής:

1. Ιδιωτικότητα Διαδικτύου (Internet Privacy): Όλα τα προσωπικά δεδομένα που διακινούνται στο διαδίκτυο είναι αφορμή για να προκύψουν θέματα που αφορούν την ιδιωτικότητα του ατόμου.
2. Ιδιωτικότητα της Οικονομίας (Financial Privacy): Τα οικονομικά δεδομένα ενός ατόμου χρήζουν επίσης προστασίας καθώς μπορεί το άτομο να πέσει θύμα οικονομικής απάτης.
3. Ιδιωτικότητα της Υγείας (Medical Privacy): Όλα τα ιατρικά αρχεία σε οργανισμούς υγείας προστατεύονται με τους πιο αυστηρούς κανόνες

πρόσβασης σε εξουσιοδοτημένα άτομα (Techopedia-
<https://www.techopedia.com/definition/10380/information-privacy>).

1.3 Η έννοια των προσωπικών δεδομένων

Η προστασία των προσωπικών δεδομένων αποτελεί ένα θεμελιώδες δικαίωμα του ανθρώπου, το οποίο κατοχυρώνεται σε πλήθος νομοθετημάτων. Στην Ευρωπαϊκή Ένωση είναι σε ισχύ η Οδηγία 95/46/EK (η οποία θα αντικατασταθεί το Μάιο του 2018 από το νέο Γενικό Κανονισμό Προστασίας Προσωπικών Δεδομένων). Στην Ελλάδα, ο νόμος που έχει ενσωματώσει την Οδηγία 95/46/EK είναι ο ν. 2472/1997. Σύμφωνα με το νόμο αυτό, προσωπικά δεδομένα (ή Δεδομένα προσωπικού χαρακτήρα) είναι κάθε πληροφορία, άμεση ή έμμεση που αναφέρεται σε φυσικό πρόσωπο και χαρακτηρίζει το υποκείμενο από φυσική, βιολογική, ψυχολογική, οικονομική, πολιτιστική ή κοινωνική άποψη.

Όταν από τα δεδομένα δεν μπορεί να εξαχθεί η ταυτότητα του προσώπου στο οποίο αναφέρονται, τότε τα δεδομένα θεωρούνται ανώνυμα δεδομένα, σε αυτήν την περίπτωση, δεν εφαρμόζεται η νομοθεσία προσωπικών δεδομένων. Αυτό επισημαίνεται ρητώς στην Οδηγία 95/46/EK, όπου αναφέρεται ότι “οι αρχές της προστασίας δεν εφαρμόζονται σε δεδομένα που έχουν καταστεί ανώνυμα κατά τρόπο ώστε να μην μπορεί να εξακριβωθεί πλέον η ταυτότητα του προσώπου στο οποίο αναφέρονται”. Ωστόσο, η απάντηση στο ερώτημα του αν πράγματι δεν μπορεί να εξακριβωθεί η ταυτότητα του προσώπου δεν είναι εύκολη: ήδη στην Οδηγία 95/46/EK γίνεται αναφορά στο ότι “για να διαπιστωθεί αν η ταυτότητα ενός προσώπου μπορεί να εξακριβωθεί, πρέπει να λαμβάνεται υπόψη το σύνολο των μέσων που μπορούν ευλόγως να χρησιμοποιηθούν, είτε από τον υπεύθυνο της επεξεργασίας, είτε από τρίτο, για να εξακριβωθεί η ταυτότητα του εν λόγω προσώπου”. Συνεπώς, με απλά λόγια, πρέπει να ληφθεί υπόψη κάθε δυνατός τρόπος και κάθε δυνατό μέσο που μπορεί να έχει κανείς, προκειμένου να αποφανθούμε αν πράγματι δεν είναι εφικτή η ταυτοποίηση προσώπων από “ανώνυμα” δεδομένα. Αυτό ακριβώς αποτελεί και τον κύριο άξονα της παρούσας διατριβής.

1.4 Δομή μεταπτυχιακής διατριβής

Η δομή της μεταπτυχιακής διατριβής είναι ως εξής:

Στο δεύτερο κεφάλαιο παρουσιάζονται παραδείγματα και περιπτώσεις επιθέσεων σε μεγάλα σύνολα δεδομένων προκειμένου να αποδειχθεί πόσο εύκολο είναι να παραβιαστεί η ιδιωτικότητα του ατόμου από μη ορθή εφαρμογή τεχνικών ανωνυμοποίησης.

Στο τρίτο κεφάλαιο παρουσιάζονται οι πιο γνωστές τεχνικές ανωνυμοποίησης που υπάρχουν και έχουν προταθεί αυτή τη στιγμή μαζί με τα πλεονεκτήματα τους αλλά και τα τρωτά τους σημεία που είναι ευάλωτα σε επιθέσεις.

Στο τέταρτο κεφάλαιο γίνεται πρακτική εφαρμογή τεχνικών ανωνυμοποίησης, εφαρμοζόμενη σε δημόσια προσβάσιμη λίστα προσωπικών δεδομένων. Στόχος της εφαρμογής αυτής είναι να καταδείξει στην πράξη τα πλεονεκτήματα αλλά και τους περιορισμούς που υπάρχουν κατά τη χρήση κλασικών τεχνικών ανωνυμοποίησης. Έμφαση θα δοθεί στα ζητήματα που ανακύπτουν από τη χρήση των τεχνικών της K - Ανωνυμίας και της L - Διαφορετικότητας (ή αλλιώς, L-ποικιλομορφίας).

Στο πέμπτο κεφάλαιο, τέλος, θα παρουσιασθούν και σχολιασθούν τα διάφορα συμπεράσματα που εξήχθησαν από το περιεχόμενο της μελέτης αυτής.

Κεφάλαιο 2

Γενικά Σχόλια

2.1 Εισαγωγή

Η ανωνυμοποίηση δεν είναι μία εύκολη υπόθεση. Υπάρχουν πάρα πολλά παραδείγματα που αποδεικνύουν ότι από μία φαινομενικά ανωνυμοποιημένη λίστα μπορεί κανείς να ανακαλύψει την ταυτότητα κάποιου προσώπου του οποίου τα δεδομένα υπάρχουν στη λίστα αυτή. Αυτό με τη σειρά του μπορεί να οδηγήσει σε αποκάλυψη σημαντικών για αυτό πληροφοριών. Ακόμα κι αν πληροφορίες όπως το όνομα, μοναδικά αναγνωριστικά (π.χ. αριθμός ταυτότητας, ΑΦΜ), διεύθυνση ή ο ταχυδρομικός κώδικας για ένα πρόσωπο έχουν αφαιρεθεί, ο επιτιθέμενος αξιοποιώντας κάποια γενικότερη γνώση που διαθέτει ή ακόμα και πρόσθετες πληροφορίες από κάποια άλλη διαθέσιμη βάση δεδομένων μπορεί να ταυτοποιήσει κάποιο άτομο στο «ανωνυμοποιημένο» σύνολο των δεδομένων. Το κύριο πρόβλημα μίας διαδικασίας ανωνυμοποίησης είναι το γεγονός ότι στην ανωνυμοποιημένη λίστα πρέπει να εμφανίζονται πολλές πληροφορίες, ακριβώς γιατί όσο περισσότερα είναι τα διαθέσιμα δεδομένα τόσο ευχερέστερη είναι π.χ. η επιστημονική ανάλυση αυτών: από την άλλη πλευρά όμως, τόσο μεγαλώνει το ενδεχόμενο να αποκαλυφθεί η ταυτότητα κάποιου προσώπου και, άρα, η λίστα να μην είναι πλήρως ανωνυμοποιημένη (Alexin, 2014).

2.2 Περιστατικά Επιθέσεων και κακής ανωνυμοποίησης

Σε αυτήν την ενότητα αναλύονται γνωστά περιστατικά επιθέσεων σε ανωνυμοποιημένους πίνακες, όπως είναι αυτό της NetFlix και της AOL καθώς επίσης και άλλα περιστατικά κακής ανωνυμοποίησης.

2.2.1 Το περιστατικό της NetFlix

Το περιστατικό της NetFlix αποτελεί ένα χαρακτηριστικό παράδειγμα στο γεγονός ότι η ανωνυμοποίηση δεν είναι πανάκεια ούτε προσδίδει 100% ασφάλεια στην προστασία ευαίσθητων προσωπικών δεδομένων. Αυτό αποδείχθηκε από μία ομάδα ερευνητών από το πανεπιστήμιο του Austin του Τέξας σε ένα πείραμά που πραγματοποίησαν το 2008. Αφορμή για την ερευνά τους στάθηκε ένας διαγωνισμός απονομής βραβείων της NetFlix. Ο διαγωνισμός βασιζόταν σε έναν αλγόριθμο ο οποίος δεχόταν κριτικές και βαθμολογίες ταινιών από τους χρήστες (Narayanan & Shmatikov, 2008).

Η NetFlix αργότερα έδωσε στην δημοσιότητα έναν ανωνυμοποιημένο πίνακα των 100.480.507 κριτικών για 17.770 ταινίες που πραγματοποιήθηκαν από 480.189 συνδρομητές για το διάστημα από τον Δεκέμβριο του 1999 έως και τον Δεκέμβριο του 2005 που περιελάμβανε πέντε πεδία. Τον τίτλο της ταινίας, της βαθμολογία της ταινίας, την ημερομηνία και την ώρα που έγινε η βαθμολογία καθώς και ένα αναγνωριστικό - τυχαίος κωδικός - για κάθε χρήστη.

Η ανωνυμοποίηση που είχε πραγματοποιηθεί στον πίνακα αυτόν ήταν ελλιπής, καθώς είχαν ανωνυμοποιηθεί τα ονόματα χρηστών, ωστόσο τα υπόλοιπα πεδία δεν ήταν ανωνυμοποιημένα - κατά την έννοια που θα περιγράψουμε στη συνέχεια της παρούσας διατριβής. Αυτό όμως ήταν και το πιο ευπαθές σημείο το οποίο και εκμεταλλεύθηκαν οι ερευνητές για να πραγματοποιήσουν την επίθεσή τους.

Ο τρόπος με τον οποίο οι ερευνητές πραγματοποίησαν την επίθεσή τους ήταν πολύ απλός: χρησιμοποιώντας έναν πίνακα δεδομένων από μία τρίτη δημόσια προσβάσιμη πηγή που συγκρίθηκε με αυτόν της Netflix. Έτσι λοιπόν χρησιμοποίησαν την Διαδικτυακή Βάση Δεδομένων Ταινιών (IMDb-Internet Movie Database), η οποία είναι μία από τις μεγαλύτερες και δημοφιλείς βάσεις δεδομένων που υπάρχουν στο διαδίκτυο. Συνέκριναν πληροφορίες όπως βαθμολογίες, ώρα, ημερομηνία βαθμολόγησης και τίτλος ταινίας με την μόνη διαφορά που υπήρχε ανάμεσα στους δύο πίνακες να είναι το πεδίο με το όνομα του χρήστη, το οποίο στον πίνακα του IMDb ήταν εμφανές και ήταν ένα καλό στοιχείο που θα οδηγούσε τους ερευνητές σε πιθανή ταυτοποίηση. Εάν ανάμεσα σε δύο εγγραφές υπήρχε ακριβής ταύτιση των τιμών των χαρακτηριστικών, τότε βγαίνει το συμπέρασμα ότι πολύ πιθανόν ένας χρήστης που είχε βαθμολογήσει στην ιστοσελίδα του Netflix να είχε βαθμολογήσει επίσης και στην ιστοσελίδα του IMDb.

Για να συγκρίνουν τους δύο πίνακες δημιούργησαν μία μετρική ομοιότητας σύγκρινε τις τιμές του χρησιμοποιώντας το περιθώριο σφάλματος. Έτσι είχαν την δυνατότητα να συμπεράνουν εάν οι δύο εγγραφές άνηκαν στον ίδιο χρήστη (M Fung, Wang, Chen, Yu, & M Fung, 2010).

2.2.2 Το περιστατικό της AOL

Το περιστατικό της AOL αφορά την αναγνώριση μιας εγγραφής σε ένα σύνολο δεδομένων που είχε δημοσιεύσει η ίδια η εταιρεία για ερευνητικούς σκοπούς. Συγκεκριμένα το σύνολο δεδομένων περιελάμβανε εγγραφές με ερωτήματα χρηστών (queries) που είχαν πραγματοποιηθεί στο διάστημα από την 1η Μαρτίου του 2006 έως 31 Μαΐου του 2006. Το σύνολο δεδομένων περιείχε 20 εκατομμύρια ερωτήματα που τέθηκαν από 650.000 χρήστες της AOL (Singh, Philosophy, & Science, 2012) .

Προτού η AOL δημοσιεύσει τον πίνακα αντικατέστησε το πεδίο του ονόματος με ένα τυχαίο αναγνωριστικό ID. Ωστόσο η New York Times εξετάζοντας συγκεκριμένους όρους ερωτημάτων μπόρεσε να ταυτοποιήσει αρκετούς από αυτή τη λίστα: χαρακτηριστικό είναι το παράδειγμα του αναγνωριστικού με τον αριθμό 4417749 το οποίο βρέθηκε ότι άνηκε στην

Thelma Arnold, μια χήρα εξήντα δύο ετών από την περιοχή του Lilburn στην πολιτεία της Georgia (Singh et al., 2012) .



Εικόνα 1: Η Thelma Arnold

(Πηγή:<http://www.nytimes.com/2006/08/09/technology/09aol.html>)

Η ταυτοποίηση ήταν σχετικά εύκολη για τους δημοσιογράφους των New York Times καθώς είχαν στην διάθεσή τους εκατοντάδες ερωτήματα με αναζητήσεις που είχε πραγματοποιήσει η Thelma Arnold σε διάστημα τριών μηνών. Αναζητήσεις με περιεχόμενο όπως το όνομα “Arnold”, “Αρχιτέκτονες τοπίου στο Lilburn” ή αναζητήσεις για πωλήσεις σπιτιών στην περιοχή της Σκοτεινής Λίμνης του Lilburn ήταν περιπτώσεις που οδήγησαν τους δημοσιογράφους στην εξηνταδύαχρονη χήρα (Narayanan & Shmatikov, 2008) .

Η AOL μετά από αυτό το συμβάν αφαίρεσε όλα αυτά τα δεδομένα από την ιστοσελίδα της σε διάστημα μίας εβδομάδας, απολογούμενη για αυτό που συνέβη.

2.2.4 Genomic Data Re-identification

Το Personal Genomic Project είναι ένα έργο το οποίο ξεκίνησε το 2005 από το Πανεπιστήμιο του Χάρβαρντ όπου σε αυτή την έρευνα οι εθελοντές που

πήραν μέρος πρόσφεραν πληροφορίες για το DNA τους, για την συμπεριφορά τους, για την κατάσταση της υγείας τους, για τα φυσικά χαρακτηριστικά τους, δημογραφικές πληροφορίες και άλλα πολλά. Το όφελος από όλες αυτές τις πληροφορίες θα βοηθούσε τους επιστήμονες να αναλύσουν τις συσχετίσεις που παρουσιάζονται ανάμεσα σε αυτά τα χαρακτηριστικά των ατόμων (Singh et al., 2012).

Το ερευνητικό έργο δίνει την ευχέρεια στον κάθε χρήστη να επιλέξει την ποσότητα της πληροφορίας την οποία θέλει να δημοσιευθεί. Συνεπώς ο χρήστης επιλέγει αν ανάμεσα στα δημογραφικά του στοιχεία για παράδειγμα, θα περιλαμβάνονται πληροφορίες όπως ο ταχυδρομικός κώδικας, το φύλο, η ημερομηνία γέννησης κ.α. Η ανωνυμοποίηση πραγματοποιήθηκε αντικαθιστώντας απλά τα ονόματα των χρηστών με έναν αναγνωριστικό αριθμό.

Αυτός ο τύπος ανωνυμοποίησης αποδεικνύει όμως με τον καλύτερο τρόπο ότι δεν είναι ασφαλής καθώς ακόμα κι αν το όνομα του χρήστη έχει αντικατασταθεί με έναν αριθμό, είναι εφικτό να γίνει η αναγνώριση ενός ατόμου από τα υπόλοιπα χαρακτηριστικά του. Αυτό απέδειξε η Latanya Sweeney.

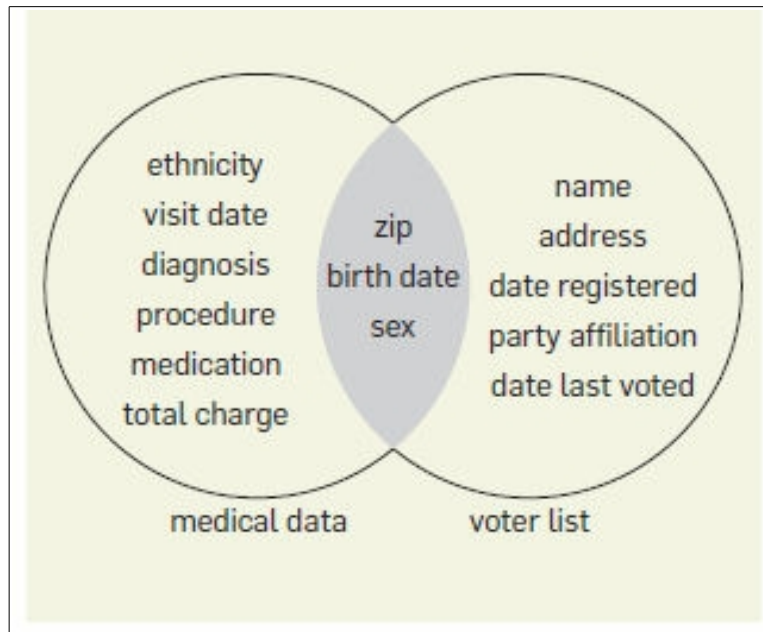
Η απόδειξη της Latanya Sweeney στηρίχθηκε σε πιο παλιά της έρευνα που ανέφερε ότι το 87% των Αμερικανών πολιτών αναγνωρίζεται μοναδικά από τον συνδυασμό των χαρακτηριστικών, φύλο, ημερομηνία γέννησης και ταχυδρομικός κώδικας. Στηριζόμενη λοιπόν στους εκλογικούς καταλόγους που ήταν διαθέσιμοι αλλά και σε άλλα σύνολα δεδομένων επίσης δημόσια κατάφερε να πραγματοποιήσει ταυτοποίηση. Αρχικά παρατήρησε ότι από τις 1130 εγγραφές που είχε το σύνολο δεδομένων του ως άνω προγράμματος οι 579 είχαν τα πλήρη στοιχεία τους εμφανή, όπως για παράδειγμα την ημερομηνία γέννησης, τον ταχυδρομικό κώδικα και και το φύλο. Επιπλέον με την βοήθεια των συνόλων δεδομένων κατέληξε ότι από τις 579 εγγραφές υπήρξε πλήρη ταύτιση με τις 241.

Τα αποτελέσματα ήταν ιδιαίτερα ακριβή γεγονός που διαπιστώθηκε και από την ερευνητική ομάδα του εν λόγω προγράμματος επιβεβαιώνοντας ότι το 84% των εγγραφών είχε ταυτοποιηθεί σωστά.

2.2.5 Η περίπτωση του κυβερνήτη της Μασαχουσέτης William Weld

Ο Γουίλιαμ Γουέλντ (William Weld) ήταν από το 1991 έως το 1997 κυβερνήτης της πολιτείας της Μασαχουσέτης. Ένα σοβαρό περιστατικό που είχε με την υγεία του στάθηκε η αφορμή για να αποδειχθούν οι αδυναμίες που έχουν οι ανωνυμοποιημένοι πίνακες που δημοσιεύονται από διάφορους οργανισμούς. Όλα ξεκίνησαν όταν ο κυβερνήτης Γουίλιαμ Γουέλντ, κατά της διάρκειας μιας διάλεξης του στο κολλέγιο Bentley, αισθάνθηκε μία αδιαθεσία και κατέρρευσε και διακομίσθηκε στο νοσοκομείο Deaconess-Waltham. Η GIC (Group Insurance Commission) αργότερα δημοσίευσε ανωνυμοποιημένα δεδομένα με βασικό σκοπό της, την βελτίωση της ιατροφαρμακευτικής περίθαλψης και τον έλεγχο του κόστους. Από το σύνολο των δεδομένων είχαν αφαιρεθεί-κατά την κρίση της GIC- όλα τα αναγνωριστικά τα οποία θα έβαζαν σε κίνδυνο την ιδιωτικότητα των ατόμων που βρίσκονταν μέσα σε αυτό. Παρόλα αυτά η Latanya Sweeney που ήταν τότε μεταπτυχιακή φοιτήτρια στο MIT (Massachusetts Institute of Technology) κατάφερε να αναγνωρίσει την εγγραφή του κυβερνήτη Γουίλιαμ Γουέλντ που βρισκόταν μέσα σε αυτό το σύνολο δεδομένων.

Ο κυβερνήτης Γουέλντ ήταν κάτοικος του Κέμπριτζ, μία πόλη των εκατό χιλιάδων κατοίκων με επτά ταχυδρομικούς κωδικούς. Μάλιστα, ο μισός πληθυσμός της πόλης είναι εγγεγραμμένος στους εκλογικούς καταλόγους. Οι εκλογικοί κατάλογοι περιλαμβάνουν πληροφορίες όπως το όνομα, ταχυδρομικό κώδικα, ημερομηνία γέννησης και το φύλο του κάθε ψηφοφόρου. Πληροφορίες που ήταν αρκετές για την Sweeney για να κάνει την ταυτοποίηση.



Εικόνα 2: Συσχέτιση δύο διαφορετικών πινάκων

Πηγή:<http://cacm.acm.org/magazines/2015/12/194640-how-to-de-identify-your-data/abstract>

Μετά από σύγκριση λοιπόν που πραγματοποίησε μεταξύ του πίνακα του GIC και των εκλογικών καταλόγων έβγαλε τα εξής συμπεράσματα. Μόλις έξι άτομα είχαν την ίδια ημερομηνία γέννησης με τον κυβερνήτη. Τρία άτομα ήταν αρσενικού γένους και ενώ ένα άτομο βρέθηκε να έχει τον ίδιο ταχυδρομικό κωδικό. Συνεπώς κατέληξε ότι η εγγραφή του κυβερνήτη βρισκόταν στο σύνολο δεδομένων του GIC (Singh et al., 2012) .

2.2.6 Το πείραμα του χρήστη X

Το πείραμα του χρήστη X πραγματοποιήθηκε από μία ομάδα ερευνητών η οποία απέδειξε πως μπορεί κάποιος έχοντας κάποιες κατάλληλες πληροφορίες να καταλήξει στα χαρακτηριστικά και εν συνεχεία στην ταυτοποίηση του ίδιου του ατόμου. Ο χρήστης X είναι ένα τυχαίο άτομο που επιλέχθηκε τυχαία από το σύνολο δεδομένων της AOL.

Οι ερευνητές μελετώντας την δραστηριότητα που είχε ο χρήστης στο ίντερνετ και το ιστορικό των αναζητήσεων του κατάφεραν να συλλέξουν πολύ σημαντικές πληροφορίες γι' αυτόν. Μερικές από αυτές τις

πληροφορίες ήταν διάφορες διευθύνσεις, ένα όνομα, προτιμήσεις σε τηλεοπτικές εκπομπές και πληροφορίες που είχαν σχέση με κάποια χόμπι που ίσως ο χρήστης X να ασχολείται.

Επιπλέον ανακάλυψαν διευθύνσεις email οι οποίες είχαν το ίδιο όνομα χρήστη με διαφορετικό πάροχο, για παράδειγμα X---@gmail.com ή X---@yahoo.com. Ακόμα βρέθηκαν διάφορες δημοσιεύσεις που είχε κάνει ο χρήστης σε διάφορα φόρουμ, τα έξι πρώτα ψηφία από ενός τηλεφωνικού αριθμού καθώς επίσης την ηλικία του και το φύλο του.

Αυτή η επίθεση που πραγματοποιήθηκε το 2013, αποδεικνύει ότι προσωπικά στοιχεία τα οποία είναι δημοσιευμένα στο διαδίκτυο, μπορούν εύκολα να ανακαλυφθούν και να οδηγήσουν σε ταυτοποίηση ενός ατόμου και με αυτόν το τρόπο να παραβιαστεί η ιδιωτικότητά του (Singh et al., 2012).

2.4 Πόσο δύσκολη είναι η ανωνυμοποίηση

Η ανωνυμοποίηση σαν τεχνική είναι συνυφασμένη με την ιδιωτικότητα των προσωπικών δεδομένων. Όταν δίνεται στην δημοσιότητα ένα σύνολο δεδομένων πρέπει το ακριβές αναγνωριστικό, όπως όνομα ή διεύθυνση κατοικίας που ουσιαστικά αναγνωρίζουν το άτομο να ανωνυμοποιούνται. Η δυσκολία εφαρμογής της ανωνυμοποίησης έγκειται στο γεγονός ότι αφενός πρέπει σε ένα σύνολο δεδομένων που δημοσιεύεται να μην είναι αναγνωρίσιμα τα πρόσωπα που βρίσκονται μέσα σε αυτό και αφετέρου, αυτό το σύνολο δεδομένων να έχει υποστεί ένα ποσοστό ανωνυμοποίησης τέτοιο ώστε, να είναι χρήσιμο για στατιστική ανάλυση και εξαγωγή χρήσιμων συμπερασμάτων και όχι να χαθεί κάθε χρησιμότητα των δεδομένων.

2.5 Απώλεια Πληροφορίας

Οι στόχος των τεχνικών ανωνυμοποίησης εκτός από το να προστατεύουν την ιδιωτικότητα σε έναν πίνακα δεδομένων, είναι επίσης να διατηρούν και την αξία των δεδομένων και της πληροφορίας σε τέτοιο επίπεδο που να είναι χρήσιμα για περαιτέρω στατιστική ανάλυση. Η απώλεια της

πληροφορίας προκύπτει όταν κατά την διάρκεια της ανωνυμοποίησης οι αρχικές τιμές των δεδομένων παραποιούνται που αυτό έχει σαν συνέπεια να χάνουν την αρχική τους αξία και παράλληλα και την χρησιμότητά τους.

Υπάρχει πληθώρα μετρικών που υπολογίζουν την αξία και την χρησιμότητα των δεδομένων κατόπιν της ανωνυμοποίησης. Μία από αυτές τις μετρικές, είναι η μετρική “Απώλεια Πληροφορίας” (ILoss) η οποία υπολογίζει την ποιότητα των δεδομένων σε έναν ανωνυμοποιημένο πίνακα και την συγκρίνει με την ποιότητα των δεδομένων στον πρωτότυπο πίνακα.

Η Απώλεια Πληροφορίας προτάθηκε από τους Xiao και Tao με σκοπό να υπολογίζει την απώλεια της πληροφορίας που προκύπτει κατά την μετατροπή μιας τιμής ενός χαρακτηριστικού από ειδική σε μία πιο γενική μορφή. Η μετρική Απώλεια Πληροφορίας προκύπτει από τον ακόλουθο τύπο:

$$ILoss(v_g) = \frac{|Vg|-1}{|DA|}$$
 όπου Vg οι αρχικές τιμές του πίνακα και D_A και γενικευμένες τιμές του πίνακα.

2.6 Γενίκευση (Generalization)

Η γενίκευση είναι μια από τις βασικές μεθόδους που χρησιμοποιείται στην τεχνική της K-Ανωνυμίας. Η έννοια της γενίκευσης ουσιαστικά αφορά την αντικατάσταση μιας τιμής ενός χαρακτηριστικού, σε μία άλλη, πιο γενικευμένη μορφή. Η γενίκευση μπορεί να εφαρμοστεί και σε αριθμητικά και σε μη αριθμητικά (κατηγορικά) δεδομένα. Πιο συγκεκριμένα, σε ένα αριθμητικό πεδίο, όπως είναι ο ταχυδρομικός κώδικας, για τις ανάγκες της γενίκευσης μπορεί να αφαιρεθεί το τελευταίο ψηφίο από τα δεξιά. Έτσι αν υποτεθεί ότι σε έναν πίνακα, στο πεδίο του ταχυδρομικού κωδικού υπάρχουν οι τιμές: {17236, 17237, 11741, 11742, 27100, 27103}, τότε μετά την γενίκευση προκύπτουν οι τιμές: {1723*, 1174*, 2710*}. Στον παρακάτω πίνακα παρουσιάζεται ακριβώς η μορφή των τιμών που μπορούν να προκύψουν ύστερα από την απαλοιφή του τελευταίου ψηφίου.

Ημ. Γέννησης	Φύλο	Τ. Κ	Ασθένεια
26/12/1965	Άνδρας	1723*	Γρίπη
02/11/1958	Άνδρας	1723*	Εγκεφαλικό
12/06/1985	Γυναίκα	1174*	Σπάσιμο χεριού
05/02/1950	Γυναίκα	1174*	Εγκεφαλικό
11/12/1945	Γυναίκα	1174*	Καρκίνος
10/10/1960	Γυναίκα	2710*	Καρκίνος
30/05/1956	Άνδρας	2710*	Διαβήτης
17/06/1940	Άνδρας	2710*	Διαβήτης
15/03/1963	Άνδρας	2710*	Εγκεφαλικό

Πίνακας 1: Γενικευμένο Αριθμητικό Πεδίο

2.7 Κατάργηση (Suppression)

Η γενίκευση όπως αναφέρθηκε παραπάνω είναι μία μέθοδος όπου μετατρέπει την τιμή ενός ψευδοαναγνωριστικού σε μία πιο γενική μορφή. Η μέθοδος της κατάργησης ουσιαστικά έρχεται και εφάπτεται με την μέθοδο της γενίκευσης για την καλύτερη δυνατή ανωνυμοποίηση. Πιο συγκεκριμένα με την μέθοδο της κατάργησης, αφαιρούνται δεδομένα όπου μετά την εφαρμογή της γενίκευσης θεωρούνται “ακραίες” τιμές (outliers) (Samarati, 2001a).

Παρακάτω με την βοήθεια των πινάκων παρατηρείται πως εφαρμόζεται η κατάργηση σε μία εγγραφή η οποία θεωρείται ως ακραία τιμή. Πιο συγκεκριμένα υπάρχει ο Πίνακας 2 ο οποίος αποτελείται από τέσσερα πεδία τα οποία είναι η ηλικία, το φύλο, ο ταχυδρομικός κώδικας και η ασθένεια του ατόμου. Στο πίνακα αυτόν έχει εφαρμοστεί η τεχνική της ανωνυμοποίησης για $k=2$ και η μέθοδος της γενίκευσης στα πεδία της ηλικίας και του ταχυδρομικού κώδικα. Από την παραπάνω διαδικασία προέκυψε ο Πίνακας 3 και όπως φαίνεται στο πεδίο της ηλικίας έχουν σχηματιστεί δύο κλάσεις ισοδυναμίας που καλύπτουν το εύρος ηλικιών από 60-75 έτη και από 45-60 έτη. Επιπλέον στο πεδίο του ταχυδρομικού κώδικα έγινε απαλοιφή του τελευταίου ψηφίου και αντικαταστάθηκε με το σύμβολο “*”. Όπως παρατηρείται στον Πίνακα 3, η τελευταία εγγραφή δεν ανήκει σε καμία κλάση ισοδυναμίας. Συνεπώς χαρακτηρίζεται ως μοναδική και κατ’

επέκταση θεωρείται πως είναι μία ακραία τιμή. Άρα σε αυτή την περίπτωση εφαρμόζεται η μέθοδος της κατάργησης και απαλείφεται όλη η εγγραφή. Οπότε με αυτόν τον τρόπο ομαλοποιείται ακόμα περισσότερο η διαδικασία της ανωνυμοποίησης. Το τελικό αποτέλεσμα που προκύπτει αποτυπώνεται στον Πίνακα 4 όπου είναι ένας πίνακας έξι εγγραφών πλέον, πλήρως ανωνυμοποιημένος.

Οι μέθοδοι της Γενίκευσης και της Κατάργησης είναι δύο διαφορετικές προσεγγίσεις που πρέπει να εφαρμόζονται κατά την διαδικασία της ανωνυμοποίησης παράλληλα για την επίτευξη του καλύτερου δυνατού αποτελέσματος. Προφανώς από την μία, σε περίπτωση που εφαρμοστεί μόνο η μέθοδος της Γενίκευσης σε έναν πίνακα τότε πολύ πιθανόν να προκύψουν πολλές μοναδικές εγγραφές, ενώ από την άλλη, αν σε περίπτωση που εφαρμοστεί μόνο η μέθοδος της κατάργησης τότε θα πρέπει όλες οι εγγραφές του πίνακα να διαγραφούν (Samarati, 2001a).

Ηλικία	Φύλο	Τ. Κ	Ασθένεια
61	Άνδρας	17237	Σπάσιμο χεριού
75	Άνδρας	17236	Σπάσιμο χεριού
66	Γυναίκα	17241	Ηπατίτιδα
56	Γυναίκα	17242	Ηπατίτιδα
51	Άνδρας	17237	AIDS
46	Γυναίκα	17223	Ηπατίτιδα
47	Άνδρας	17238	AIDS

Πίνακας 2: Νοσοκομειακή Λίστα Ασθενών

Ηλικία	Φύλο	Τ. Κ	Ασθένεια
60-75	Άνδρας	1723*	Σπάσιμο χεριού
60-75	Άνδρας	1723*	Σπάσιμο χεριού
45-60	Άνδρας	1723*	AIDS
45-60	Άνδρας	1723*	AIDS
45-60	Γυναίκα	1724*	Ηπατίτιδα
45-60	Γυναίκα	1724*	Ηπατίτιδα
45-60	Γυναίκα	1722*	Ηπατίτιδα

Πίνακας 3: Ανωνυμοποιημένος Πίνακας με $k=2$

Ηλικία	Φύλο	Τ. Κ	Ασθένεια
60-75	Άνδρας	1723*	Σπάσιμο χεριού
60-75	Άνδρας	1723*	Σπάσιμο χεριού
45-60	Άνδρας	1723*	AIDS
45-60	Άνδρας	1723*	AIDS
45-60	Γυναίκα	1724*	Ηπατίτιδα
45-60	Γυναίκα	1724*	Ηπατίτιδα

Πίνακας 4: Πίνακας που έχει εφαρμοστεί η Κατάργηση

Κεφάλαιο 3

Τεχνικές Ανωνυμοποίησης

3.1 Εισαγωγή

Στην παρούσα διατριβή μελετώνται τεχνικές ανωνυμοποίησης οι οποίες εφαρμόζονται σε πίνακες δεδομένων: κάθε πίνακας αποτελείται από διάφορες εγγραφές (γραμμές του πίνακα), όπου κάθε εγγραφή αντιστοιχεί σε ένα πρόσωπο. Κάθε εγγραφή λαμβάνει τιμές σε ένα σύνολο γνωρισμάτων/χαρακτηριστικών, όπως ονοματεπώνυμο, διεύθυνση, ημερομηνία γέννησης κτλ. (στήλες του πίνακα).

Τα χαρακτηριστικά σε έναν πίνακα δεδομένων χωρίζονται σε τρεις κατηγορίες. Τα **ευαίσθητα χαρακτηριστικά** (SA-Sensitive Attribute), τα **ψευδοαναγνωριστικά** (QI-Quasi Identifier) και τα **ακριβή αναγνωριστικά** (EI-Explicit Identifier).

Το ευαίσθητο χαρακτηριστικό αναφέρεται στο χαρακτηριστικό το οποίο περιέχει σημαντικές για το άτομο πληροφορίες και πρέπει να προστατευθεί. Τα ψευδοαναγνωριστικά είναι χαρακτηριστικά τα οποία από μόνα τους δεν επαρκούν για τη μοναδική ταυτοποίηση του ατόμου αλλά, εάν συγκριθούν με κάποιον άλλο εξωτερικό πίνακα, μπορούν ενδεχομένως να οδηγήσουν στην αναγνώριση ενός ατόμου. Ψευδοαναγνωριστικά θεωρούνται τα χαρακτηριστικά όπως ο Ταχυδρομικός Κώδικας, το φύλο, η ηλικία, το επάγγελμα ή η εθνικότητα. Τα ακριβή αναγνωριστικά αφορούν τα χαρακτηριστικά εκείνα από τα οποία είναι εφικτό να αναγνωριστεί πλήρως η ταυτότητα ενός ατόμου όπως για παράδειγμα το όνομα, ο αριθμός ταυτότητας ή ο αριθμός φορολογικού μητρώου του ατόμου.

Είναι προφανές ότι, ως πρώτο βήμα μίας διαδικασίας ανωνυμοποίησης, θα πρέπει να είναι η απομάκρυνση όλων των αναγνωριστικών. Ωστόσο, όπως

είδαμε στο Κεφάλαιο 2, αυτό δεν είναι αρκετό: ο λόγος είναι ότι και τα ψευδο-αναγνωριστικά φέρουν χρήσιμη πληροφορία που μπορεί να οδηγήσει σε ταυτοποίηση. Στη συνέχεια, αναφερόμαστε στην έννοια της κλάσης ισοδυναμίας ενός συνόλου δεδομένων ως το σύνολο των εγγραφών οι οποίες έχουν τις ίδιες τιμές στα ψευδοαναγνωριστικά (Alexin, 2014).

3.2 Κ-Ανωνυμία (K-Anonymity)

Πάρα πολλοί οργανισμοί, όπως αναφέρθηκε νωρίτερα, δημοσιεύουν πίνακες δεδομένων με σκοπό να χρησιμοποιηθούν σε έρευνες για την υγεία αλλά και για δημογραφικές έρευνες. Στις περισσότερες περιπτώσεις, δεδομένα όπως το όνομα ή ο αριθμός ταυτότητας των ατόμων που εμπεριέχονται σε αυτούς τους πίνακες αφαιρούνται έτσι ώστε να μην είναι αναγνωρίσιμα από τρίτους. Με αυτόν τον τρόπο όμως, όπως είδαμε, τα δεδομένα συνεχίζουν να μην είναι ανώνυμα. Αυτό ισχύει γιατί υπάρχει ο κίνδυνος συσχέτισης αυτού του πίνακα δεδομένων με κάποιον άλλον με τα ίδια χαρακτηριστικά.

Σύμφωνα με έρευνες που έχουν πραγματοποιηθεί, το 87% του πληθυσμού των Ηνωμένων Πολιτειών της Αμερικής μπορεί να αναγνωριστεί ατομικά ο κάθε πολίτης από τρία και μόνο χαρακτηριστικά, τον ταχυδρομικό κώδικα, το φύλο και την ημερομηνία γέννησης του. Έτσι με αυτήν και μόνο την ιδιότητα συγκρίνοντας δύο πίνακες μπορεί κάποιος να ταυτοποιήσει ένα πρόσωπο μέσα σε ένα σύνολο δεδομένων (Sweeney, 2002a). Αυτή η ιδιότητα ήταν και η αιτία για να αναπτυχθεί η τεχνική της Κ-Ανωνυμίας, προς αποφυγή επιθέσεων Αποκάλυψης Ταυτότητας (Identity Disclosure).

Η Κ-Ανωνυμία είναι μία από τις πιο διαδεδομένες τεχνικές ανωνυμοποίησης που χρησιμοποιούνται για την ανωνυμοποίηση συνόλων δεδομένων. Η Κ-Ανωνυμία (K-Anonymity) είναι μια τεχνική η οποία, μέσω των μεθόδων της γενίκευσης ή της κατάργησης που χρησιμοποιεί, αποτρέπει την αναγνώριση της εγγραφής ενός ατόμου η οποία βρίσκεται σε ένα σύνολο δεδομένων.

Ο ακριβής ορισμός της Κ-Ανωνυμίας είναι ο εξής: Ένας πίνακας δεδομένων θεωρείται ότι ικανοποιεί την k-Ανωνυμία αν και μόνο αν κάθε ακολουθία

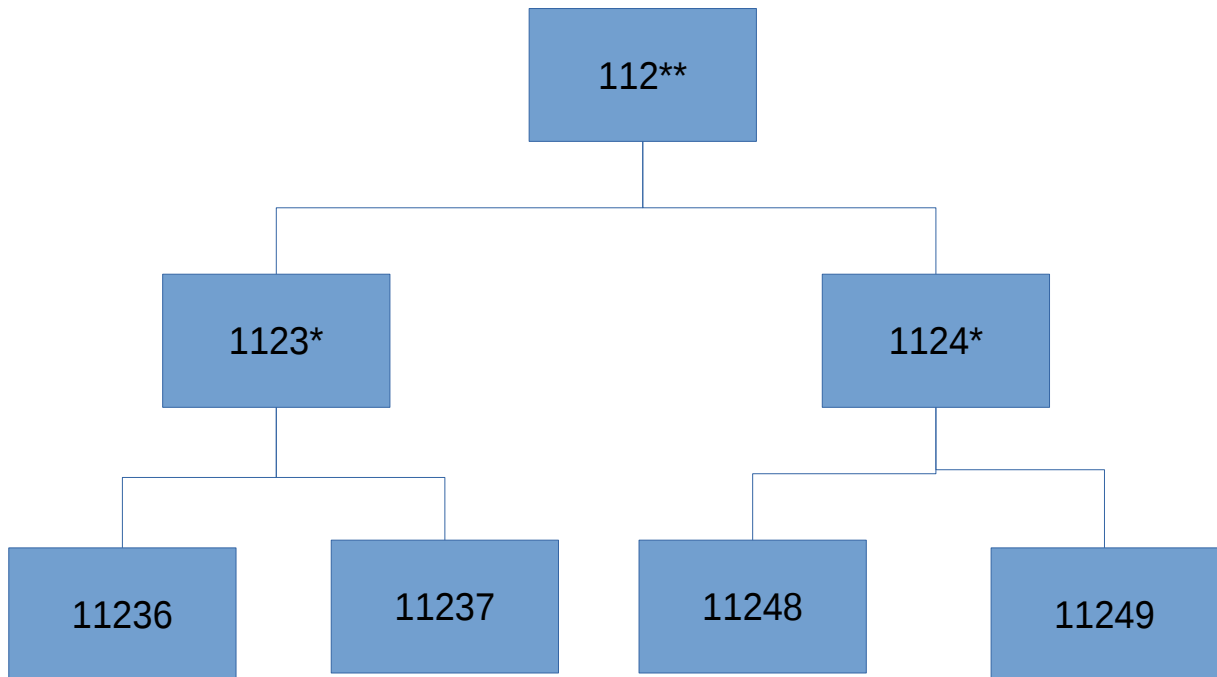
τιμών στα ψευδο-αναγνωριστικά που εμφανίζεται στον πίνακα αυτόν έχει πληθικότητα τουλάχιστον ίση με το k (Sweeney, 2002b) .

Ο παρακάτω πίνακας αποτελείται από έξι πεδία, τα τέσσερα εξ αυτών είναι τα ψευδοαναγνωριστικά (Quasi Identifier) {Φυλή, Ημ/νία Γέννησης, Φύλο και Τ. Κ} ενώ η ασθένεια είναι το ευαίσθητο χαρακτηριστικό. Στον πίνακα αυτόν έχει εφαρμοστεί k -Ανωνυμία με $k=3$ και η μέθοδος της γενίκευσης για το πεδίο του ταχυδρομικού κώδικα με απαλοιφή του τελευταίου ψηφίου. Όπως μπορεί να παρατηρηθεί υπάρχουν τουλάχιστον τρεις εγγραφές στον πίνακα όπου οι τιμές στα πεδία των ψευδοαναγνωριστικών είναι ίδιες. Αυτό συμβαίνει στις εγγραφές με Α/Α: 1, 2,3 και 11, 12, 13.

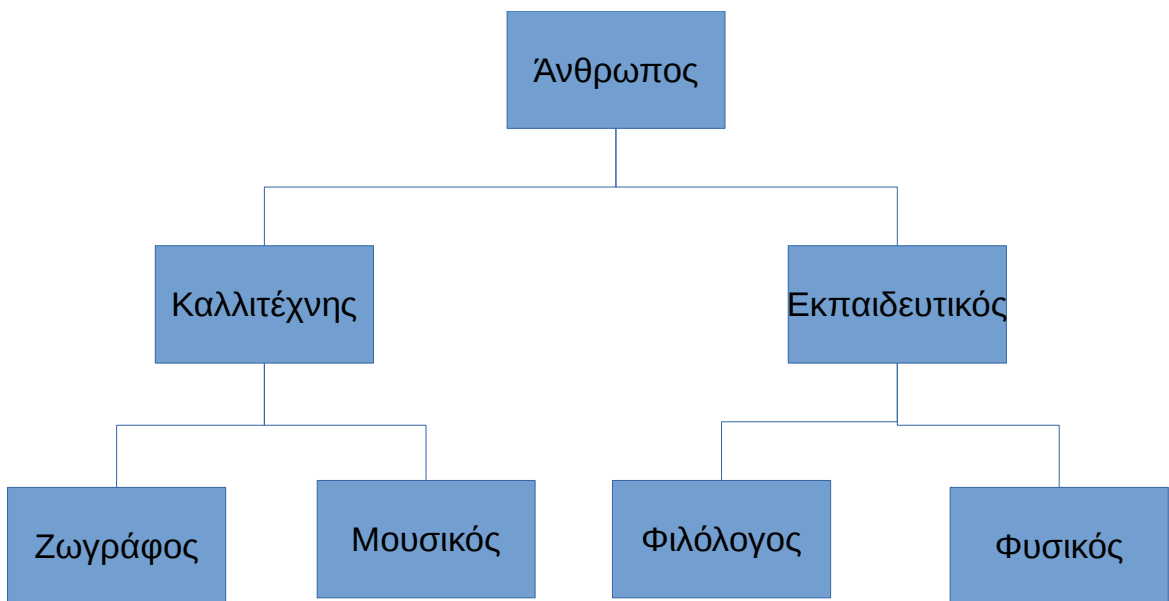
A/A	Φυλή	Ημ/νια Γέννησης	Φύλο	Τ. Κ	Ασθένεια
1	Λευκός	1965	Άνδρας	1723*	Πνευμονία
2	Λευκός	1965	Άνδρας	1723*	Βρογχίτιδα
3	Λευκός	1965	Άνδρας	1723*	Ηπατίτιδα
4	Λευκός	1970	Γυναίκα	1742*	Εγκεφαλικό
5	Λευκός	1970	Γυναίκα	1742*	Διαβήτης
6	Λευκός	1970	Γυναίκα	1742*	Πνευμονία
7	Λευκός	1970	Γυναίκα	2710*	Εγκεφαλικό
8	Μαύρος	1960	Άνδρας	2710*	Διαβήτης
9	Μαύρος	1960	Άνδρας	2710*	Βρογχίτιδα
10	Μαυρός	1960	Άνδρας	2713*	Διαβήτης
11	Μαύρος	1955	Γυναίκα	2713*	Εγκεφαλικό
12	Μαύρος	1955	Γυναίκα	2713*	Αλτσχάιμερ
13	Μαύρος	1955	Γυναίκα	2713*	Αλτσχάιμερ

Πίνακας 5: Παράδειγμα k - Ανωνυμίας με $k=3$

Φυσικά αν είναι θεμιτό μπορεί να υπάρξει ακόμα μεγαλύτερη γενίκευση στο πεδίο αποκρύπτοντας ακόμα περισσότερα ψηφία. Στην προκειμένη όμως περίπτωση, όσο μεγαλύτερη είναι η γενίκευση που γίνεται, τόσο μεγαλύτερη απώλεια πληροφορίας θα προκύψει και τα δεδομένα από ένα σημείο και μετά παύουν να είναι χρήσιμα για στατιστικές έρευνες. Στο παρακάτω διάγραμμα απεικονίζεται το δέντρο γενίκευσης ενός αριθμητικού πεδίου.



Εικόνα 3: Δέντρο Γενίκευσης Αριθμητικών Χαρακτηριστικών



Εικόνα 4: Δέντρο Γενίκευσης Κατηγορικών Χαρακτηριστικών

3.2.3 Μειονεκτήματα Κ-Ανωνυμίας

Η τεχνική της Κ-Ανωνυμίας όπως έχει ήδη αναφερθεί δημιουργήθηκε με στόχο την αντιμετώπιση των επιθέσεων αποκάλυψης ταυτότητας (Identity Disclosure). Συνεπώς, κάποιος που γνωρίζει τα ψευδοαναγνωριστικά ενός

ατόμου όπου η εγγραφή του βρίσκεται στον ανωνυμοποιημένο πίνακα, δεν μπορεί να ταυτοποιήσει την εγγραφή αυτή με εμπιστοσύνη μεγαλύτερη του $1/k$ (Li, Li, & Venkatasubramanian, 2007) .

Μπορεί όμως η τεχνική της K-Ανωνυμίας να σχεδιάστηκε για την αντιμετώπιση των επιθέσεων αποκάλυψης ταυτότητας (Identity Disclosure), παρόλα αυτά όμως φαίνεται πως δεν μπορεί να ανταποκριθεί στις Επιθέσεις Εξαγωγής Συμπεράσματος και στις Επιθέσεις Γνώσης Υπόβαθρου οι οποίες παρουσιάζονται αναλυτικότερα στην συνέχεια της ενότητας.

3.2.4 Επιθέσεις στην K-Ανωνυμία

Στην K-Ανωνυμία υπάρχουν δύο επιθέσεις, η Επίθεση Εξαγωγής Συμπεράσματος (Inference Attack ή Homogeneity Attack) και η επίθεση Γνώσης Υπόβαθρου (Background Knowledge Attack) οι οποίες παρουσιάζονται στην συνέχεια και αποδεικνύουν τα κενά ασφάλειας που έχει αυτή η τεχνική ανωνυμοποίησης.

Στην επίθεση “Εξαγωγής Συμπεράσματος” ας υποθέσει κάποιος ότι η Κέλλυ μένει στην ίδια γειτονιά με τον Γιώργο. Κάποια στιγμή ο Γιώργος διακομίζεται στο τοπικό νοσοκομείο της πόλης για λόγους υγείας. Η Κέλλυ γνωρίζοντας την ημ/νια γέννησης, τον ταχυδρομικό κώδικα και επιπλέον ότι ο Γιώργος είχε βήχα και αυξημένο πυρετό, μπορεί να καταλήξει στο συμπέρασμα ότι ο λόγος διακομιδής του στο νοσοκομείο ήταν η γρίπη όπως φαίνεται άλλωστε και στον παρακάτω πίνακα 6.

Ημ/Γέννησης	Φύλο	T. Κ	Ασθένεια
1961-1975	Άνδρας	1712*	Ηπατίτιδα
1961-1975	Άνδρας	1712*	Ηπατίτιδα
1961-1975	Άνδρας	1712*	Γρίπη
1961-1975	Άνδρας	1712*	Γρίπη
1950-1960	Γυναίκα	1174*	Σπάσιμο χεριού
1950-1960	Γυναίκα	1174*	Σπάσιμο χεριού
1950-1960	Γυναίκα	1174*	Εγκεφαλικό
1950-1960	Άνδρας	1123*	Εγκεφαλικό
1950-1960	Άνδρας	1123*	Εγκεφαλικό

Πίνακας 6: Ανωνυμοποιημένος Πίνακας με $k=2$

Η άλλη επίθεση που εμφανίζεται στην K-Ανωνυμία είναι η επίθεση της “Γνώσης Υπόβαθρου” (Background Knowledge Attack) όπου ο επιτιθέμενος έχοντας εις γνώση κάποια συγκεκριμένα δεδομένα για κάποιο άτομο μπορεί να προχωρήσει στην αποκάλυψη της ταυτότητάς του. Ας θεωρηθεί ότι ένας φίλος της Κέλλυς από την Ιαπωνία μπήκε πρόσφατα στο νοσοκομείο. Η Κέλλυ φυσικά γνωρίζει ότι μένει σε μια περιοχή με ταχυδρομικό κώδικα 17237, ότι είναι 25 χρόνων και φυσικά γνωρίζει ότι είναι άνδρας. Η Κέλλυ με αυτά τα δεδομένα ανακαλύπτει σε ποια κλάση ισοδυναμίας βρίσκεται ο φίλος της, στην οποία τα ευαίσθητα χαρακτηριστικά που υπάρχουν είναι η γρίπη και η καρδιακή προσβολή. Στην Ιαπωνία είναι γνωστό πως το ποσοστό να πάθει κάποιος καρδιακή προσβολή είναι εξαιρετικά μικρό, συνεπώς με αυτό το δεδομένο μπορεί να καταλήξει ότι ο φίλος της από την Ιαπωνία διακομίσθηκε στο νοσοκομείο επειδή είχε γρίπη.

3.3 Τεχνική L - Διαφορετικότητας (L-Diversity)

Η τεχνική Διαφορετικότητας (ή Ποικιλομορφίας) Τάξης L (L-Diversity) είναι μία ακόμα τεχνική ανωνυμοποίησης. Προτάθηκε από τον Machanavajjhala με σκοπό την αποτροπή επιθέσεων εξαγωγής συμπεράσματος. Η πιο απλή μορφή της τεχνικής L-Diversity προϋποθέτει ότι κάθε κλάση ισοδυναμίας ενός ημιαναγνωριστικού περιλαμβάνει τουλάχιστον καλά ορισμένες, διακριτές τιμές αριθμού L στο ευαίσθητο χαρακτηριστικό. Αυτή η απλή μορφή της τεχνικής Διαφορετικότητας L δεν παρέχει όμως προστασία σε

επιθέσεις εξαγωγής συμπεράσματος με την βοήθεια των πιθανοτήτων (Probabilistic Inference Attacks). Αυτό συμβαίνει γιατί εάν σε μία κλάση ισοδυναμίας ενός ψευδοαναγνωριστικού εμφανίζεται με μεγαλύτερη συχνότητα μία τιμή στο ευαίσθητο χαρακτηριστικό σε σχέση με μία άλλη τότε δίνεται η δυνατότητα στον επιτιθέμενο να εξάγει πληροφορίες για μία εγγραφή.

Στον Πίνακα έχει εφαρμοστεί η τεχνική K-Ανωνυμία με $k=2$ και η τεχνική L-Diversity με $l=2$. Όπως προκύπτει και παρακάτω με την βοήθεια του ανωνυμοποιημένου πίνακα 7, στην περίπτωση που ο επιτιθέμενος γνωρίζει ότι το άτομο που βρίσκεται στον Πίνακα είναι άνδρας, γνωρίζει επίσης την ηλικία του, τον ταχυδρομικό κώδικα της κατοικίας του ακόμα και το ιατρικό ιστορικό που μπορεί να είχε το συγκεκριμένο άτομο, τότε με 80% πιθανότητα να μαντέψει ότι αυτό το άτομο πάσχει από γρίπη. Η συγκεκριμένη ευπάθεια ήταν η αρχή για να αναπτυχθούν δύο διαφορετικές εκδόσεις της τεχνικής L-Diversity με πιο αυστηρά κριτήρια ανωνυμοποίησης, η τεχνική L - Διαφορετικότητας κατά εντροπία (entropy l-Diversity) και η Αναδρομική (Recursive) (c, l) Diversity.

Φύλο	Ηλικία	Τ. Κ	Αρρώστια
Άνδρας	60-70	271**	Γρίπη
Άνδρας	60-70	271**	Πνευμονία
Άνδρας	60-70	271**	Γρίπη
Άνδρας	60-70	271**	Πνευμονία
Άνδρας	60-70	271**	Γρίπη
Γυναίκα	45-60	374**	AIDS
Γυναίκα	45-60	374**	Ηπατίτιδα
Γυναίκα	45-60	343**	Ηπατίτιδα
Γυναίκα	45-60	343**	AIDS

Πίνακας 7: Πίνακας με L-Diversity 2

3.3.1 Οι τεχνικές Entropy L-Diversity και Recursive (c, l) Diversity

Ένα πίνακας λέγεται ότι έχει Entropy L-Diversity αν και μόνο αν σε κάθε κλάση ισοδυναμίας E ισχύει η παρακάτω ιδιότητα όπου: $Entropy(E) \geq \log(l)$. Για να υπάρχει Entropία L - Διαφορετικότητας σε κάθε κλάση ισοδυναμίας πρέπει η Entropία σε ολόκληρο τον πίνακα να είναι τουλάχιστον ίση με $\log(l)$ (Li, Li, & Venkatasubramanian, 2007).

Ο υπολογισμός της εντροπίας γίνεται από τον παρακάτω τύπο:

$$Entropy(l) = - \sum_{s \in S} p(E/s) \log p(E/s)$$

Όπου S είναι το πεδίο του ευαίσθητου χαρακτηριστικού και $p(E, s)$ είναι το κλάσμα των εγγραφών της κλάσης ισοδυναμίας E που περιέχουν το ευαίσθητο s .

Η Recursive Diversity εξασφαλίζει ότι η τιμή με την μεγαλύτερη συχνότητα δεν θα εμφανίζεται τόσο συχνά και η τιμή με την μικρότερη συχνότητα δεν θα εμφανίζεται τόσο σπάνια. Παραδείγματος χάριν, ας θεωρηθεί m ο αριθμός των τιμών του ευαίσθητου χαρακτηριστικού σε μία κλάση ισοδυναμίας και r_i όπου ισχύει $1 \leq i \leq m$, την συχνότητα εμφάνισης της i στης πιο συχνής τιμής στο ευαίσθητο χαρακτηριστικό μίας κλάσης ισοδυναμίας. Τότε μια κλάση ισοδυναμίας λέγεται ότι είναι Recursive (c, l) diverse εάν ισχύει ότι $r_1 < c (r_1 + r_{l+1} + 1 + \dots + r_m)$. Ένας πίνακας θεωρείται ότι έχει Recursive(c,l) Diversity όταν όλες οι κλάσεις ισοδυναμίας έχουν Recursive (c,l) Diversity (Li et al., 2007).

3.3.2 Μειονεκτήματα στην L - Διαφορετικότητα

Η L-Διαφορετικότητα παρόλο που παρέχει ένα επιπλέον επίπεδο προστασίας στην ανωνυμοποίηση συνόλων δεδομένων φαίνεται πως τελικά δεν είναι αποδοτική σε επιθέσεις αποκάλυψης χαρακτηριστικών (Attribute Disclosure). Επιπλέον παρουσιάζει σε μερικές περιπτώσεις περιορισμούς σε βαθμό τέτοιο που δεν μπορεί να εφαρμοστεί ή δεν κρίνεται απαραίτητο να εφαρμοστεί. Οι τύποι επιθέσεων που εμφανίζονται στην L-Διαφορετικότητα

είναι οι Επίθεση Ασυμμετρίας (Skewness Attack) και Επίθεση Ομοιότητας (Similarity Attack) και αναλύονται στη συνέχεια του κεφαλαίου.

3.3.3 Επιθέσεις της L – Διαφορετικότητας

Επίθεση Ασυμμετρίας: Ας υποθεθεί ότι υπάρχει ένας ανωνυμοποιημένος πίνακας των χιλίων εγγραφών με ένα ευαίσθητο χαρακτηριστικό το οποίο παίρνει δύο τιμές, θετικό και αρνητικό. Αυτός ο πίνακας περιέχει τις εγγραφές των ασθενών από ένα νοσοκομείο όπου το 99% των εξετάσεων που τους έγιναν είναι αρνητικό και το υπόλοιπο 1% είναι θετικό (Li et al., 2007).

Σε αυτή τη περίπτωση είναι προφανές πως υπάρχει πολύ μεγάλη διαφορά στην κατανομή των τιμών μεταξύ του αρνητικού και του θετικού, γεγονός που παρέχει μία γενική πληροφορία στο κοινό ότι πολύ μικρό ποσοστό νοσεί από τη συγκεκριμένη ασθένεια. Ωστόσο, αν αναλογιστούμε μία κλάση ισοδυναμίας όπου, π.χ. οι μισές εγγραφές έχουν την τιμή «αρνητικό» και οι άλλες μισές «θετικό» στο εν λόγω χαρακτηριστικό, τότε αν ένας επιτιθέμενος αναγνωρίζει ότι ένα πρόσωπο ανήκει σε αυτήν την κλάση ισοδυναμίας, τότε η πιθανότητά του να είναι στο χαμηλό ποσοστό του πληθυσμού που νοσεί δεν είναι τόσο μικρή: αυτό γίνεται ακόμα πιο έντονο αν, στην εν λόγω κλάση ισοδυναμίας, είναι περισσότερες οι εγγραφές με την ένδειξη «θετικό» στο εν λόγω γνώρισμα: ως συνέπεια αυτού προκύπτει θέμα παραβίασης της ιδιωτικότητας. Οπότε προκύπτει το συμπέρασμα ότι σε μια κλάση ισοδυναμίας που υπάρχουν εγγραφές όπου στο ευαίσθητο χαρακτηριστικό κυριαρχεί μόνο η μία τιμή, τότε η L-Diversity βαθμού 2 δεν είναι πλέον κατάλληλη και αποτελεσματική.

Επίθεση Ομοιότητας: Η επίθεση ομοιότητας συμβαίνει στην περίπτωση που οι τιμές του ευαίσθητου χαρακτηριστικού να μεν είναι διακριτές αλλά σημασιολογικά όμοιες. Στην περίπτωση αυτή ο επιτιθέμενος μπορεί να εξάγει χρήσιμα συμπεράσματα και μάθει σημαντικές πληροφορίες για ένα άτομο. Με βάση τους παρακάτω πίνακες μπορεί να επεξηγηθεί η παραπάνω επίθεση. Ο Πίνακας 8 είναι ο αρχικός πίνακας και ο Πίνακας 9 είναι ένας

ανωνυμοποιημένος πίνακας στον οποίο έχει εφαρμοστεί διακριτή και εντροπίας βαθμού 3 διαφορετικότητα (Samarati, 2001a).

Τα ευαίσθητα χαρακτηριστικά του πίνακα 8 είναι ο μισθός και η ασθένεια. Αν υποτεθεί ότι κάποιος γνωρίζει ότι η εγγραφή του Γιώργου βρίσκεται στην δεύτερη κλάση ισοδυναμίας του πίνακα 9, τότε αυτόματα παίρνει την πληροφορία ότι ο μισθός του Γιώργου βρίσκεται στο εύρος των 2000-4000 ευρώ. Επιπλέον η πληροφορία αυτή μπορεί να συσχετιστεί άμεσα και με το χαρακτηριστικό της ασθένειας όπου μπορεί ο επιτιθέμενος να συμπεράνει ότι ο Γιώργος έχει καρδιολογικά προβλήματα.

	Ηλικία	Τ. Κ	Μισθός	Ασθένεια
1	35	17205	1300	Γρίπη
2	26	17126	750	Βρογχίτιδα
3	55	17127	2500	Υπέρταση
4	49	17122	3500	Καρδιοπάθεια
5	24	17106	580	Πνευμονία
6	22	17117	580	Γρίπη
7	31	17217	1250	Πνευμονία
8	32	17217	1100	Βρογχίτιδα
9	60	17103	3000	Στεφανιαία Νόσος

Πίνακας 8: Στοιχεία Μισθού και Ασθένεια

	Ηλικία	Τ. Κ	Μισθός	Ασθένεια
1	2*	171**	500-1000	Γρίπη
2	2*	171**	500-1000	Βρογχίτιδα
3	2*	171**	500-1000	Πνευμονία
4	>=40	172**	2000-4000	Στεφανιαία Νόσος
5	>=40	172**	2000-4000	Καρδιοπάθεια
6	>=40	172**	2000-4000	Υπέρταση
7	3*	171**	1000-2000	Πνευμονία
8	3*	171**	1000-2000	Γρίπη
9	3*	171**	1000-2000	Βρογχίτιδα

Πίνακας 9: Ανωνυμοποιημένος Πίνακας με L-Diversity=3

3.4 T-Εγγύτητα (T-Closeness)

Η τεχνική T-Εγγύτητας είναι μία ακόμα τεχνική ανωνυμοποίησης η οποία προτάθηκε από τους Ninghui Li, Tiancheng Li και Suresh Venkatasubramanian. Ουσιαστικά η τεχνική της T-Εγγύτητας είναι μία πιο εξελιγμένη τεχνική από την Ανωθυμία Διαφορετικότητας Τάξης L και την K-Ανωθυμία, η οποία στοχεύει στην δημιουργία κλάσεων ισοδυναμίας των οποίων η κατανομή του ευαίσθητου χαρακτηριστικού τους θα πλησιάζει, ιδανικά θα ταυτίζεται, με την κατανομή του αρχικού πίνακα δεδομένων (Narayanan & Shmatikov, 2008).

Ο ακριβής ορισμός της T-Εγγύτητας αναφέρει: Μία κλάση ισοδυναμίας θεωρείται ότι έχει T-Εγγύτητα εάν η απόσταση μεταξύ της κατανομής του ευαίσθητου χαρακτηριστικού σε αυτή τη κλάση και της κατανομής του ευαίσθητου χαρακτηριστικού σε όλο τον πίνακα σε ξεπερνά ένα συγκεκριμένο κατώφλι T. Ένας πίνακας θεωρείται ότι έχει T-Εγγύτητα εάν όλες οι κλάσεις ισοδυναμίας του έχουν T-Εγγύτητα (Alexin, 2014).

3.5 Ανατομία (Anatomy)

Μία ακόμα τεχνική ανωνυμοποίησης είναι η Ανατομία. Προτάθηκε από τους Xiaokui Xiao και Yufei Tao με στόχο την διατήρηση της ιδιωτικότητας και κατ' επέκταση να μειωθεί η απώλεια πληροφορίας. Η λειτουργία της Ανατομίας βασίζεται στο να διαχωρίζει τα ψευδοαναγνωριστικά από τα ευαίσθητα χαρακτηριστικά και να αποκρύπτεται ο συσχετισμός μεταξύ τους. Πιο συγκεκριμένα δημιουργεί δύο πίνακες. Ο ένας είναι ο πίνακας με τα ψευδοαναγνωριστικά (QIT - Quasi Identifier Table) και ο άλλος πίνακας περιέχει τα ευαίσθητα χαρακτηριστικά (ST - Sensitive Table) (Xiao & Tao, 2006.).

Ο παρακάτω πίνακας 10 αποτελείται από πέντε χαρακτηριστικά τα οποία είναι: ο αύξοντας αριθμός κάθε εγγραφής, η χρονολογία γέννησης, το φύλο, ο ταχυδρομικός κώδικας και η ασθένεια του ατόμου. Τα ψευδοαναγνωριστικά του πίνακα αυτού είναι τα {Ηλικία, Φύλο, Ταχυδρομικός Κώδικας} ενώ το ευαίσθητο χαρακτηριστικό είναι η {Ασθένεια}.

ID	Ηλικία	Φύλο	Τ. Κ	Ασθένεια
1	1975	Άνδρας	17237	Γαστρεντερίτιδα
2	1964	Γυναίκα	17247	Ηπατίτιδα
3	1978	Άνδρας	17236	Ηπατίτιδα
4	1962	Γυναίκα	17246	Γρίπη
5	1961	Άνδρας	17237	Βρογχίτιδα
6	1975	Γυναίκα	17241	Γαστρεντερίτιδα
7	1975	Άνδρας	17241	Γρίπη
8	1953	Γυναίκα	17242	Ηπατίτιδα
9	1962	Άνδρας	17237	Γρίπη
10	1966	Άνδρας	17242	Γαστρεντερίτιδα

Πίνακας 10: Νοσοκομειακή Λίστα Ασθενών

Ο παραπάνω πίνακας δεν δημοσιεύεται – χρησιμοποιείται ως βοηθητικό εργαλείο, όπως περιγράφεται στη συνέχεια.

Έπειτα δημιουργείται ο παρακάτω πίνακας 11 που ικανοποιεί τα κριτήρια της K-Ανωνυμίας με $k=4$ και L- Διαφορετικότητα και με $l=3$.

ID	Ηλικία	Φύλο	Τ. Κ	Ασθένεια
1	>1965	Γυναίκα	1724*	Ηπατίτιδα
2	>1965	Γυναίκα	1724*	Ηπατίτιδα
3	>1965	Γυναίκα	1724*	Γρίπη
4	>1965	Γυναίκα	1724*	Γαστρεντερίτιδα
5	>1965	Άνδρας	1723*	Γαστρεντερίτιδα
6	>1965	Άνδρας	1723*	Ηπατίτιδα
7	<1975	Άνδρας	1723*	Βρογχίτιδα
8	<1975	Άνδρας	1723*	Γρίπη
9	<1975	Άνδρας	1723*	Γρίπη
10	<1975	Άνδρας	1723*	Γαστρεντερίτιδα

Πίνακας 11: Πίνακας με k-Ανωνυμία=4 και L-Διαφορετικότητα=3

Στην συνέχεια δημιουργείται ο πίνακας 12 με τις αρχικές τιμές των χαρακτηριστικών και την αντίστοιχη κλάση ισοδυναμίας που ανήκει η κάθε εγγραφή.

ID	Ηλικία	Φύλο	T. Κ	Κλάση Ισοδυναμίας
1	1975	Άνδρας	17237	2
2	1964	Γυναίκα	17247	1
3	1978	Άνδρας	17236	2
4	1962	Γυναίκα	17246	1
5	1961	Άνδρας	17237	2
6	1975	Γυναίκα	17241	1
7	1975	Άνδρας	17241	2
8	1953	Γυναίκα	17242	1
9	1962	Άνδρας	17237	2
10	1966	Άνδρας	17242	2

Πίνακας 12: Πίνακας με το ID της κλάσης ισοδυναμίας

Στο τέλος δημιουργείται ο παρακάτω πίνακας 13 όπου απεικονίζονται οι κλάσεις ισοδυναμίας με το ευαίσθητο χαρακτηριστικό αλλά και τον αριθμό των εμφανίσεων του ευαίσθητου χαρακτηριστικού. Αυτοί οι δύο πίνακες (12 και 13) είναι αυτοί που τελικά δημοσιοποιούνται, ως ανώνυμοι.

Στην Ανατομία μπορεί να είναι εμφανή τα ψευδοαναγνωριστικά αλλά και το ευαίσθητο χαρακτηριστικό αλλά ο επιτιθέμενος δεν μπορεί με απόλυτη σιγουριά να συσχετίσει την εγγραφή ενός ατόμου με το αντίστοιχο ευαίσθητο χαρακτηριστικό. Το κύριο πλεονέκτημα της ανατομίας είναι ότι έχουμε τη μικρότερη απώλεια πληροφορίας (αφού στον Πίνακα 12 δεν έχει γίνει γενίκευση των ψευδο-αναγνωριστικών).

Η Ανατομία χρησιμοποιείται συνήθως και προτιμάται από την γενίκευση σε περιπτώσεις όπου ο επιτιθέμενος γνωρίζει ότι το άτομο που ψάχνει είναι μέσα στον πίνακα και επίσης μπορεί να γνωρίζει και τα ψευδοαναγνωριστικά του (Xiao & Tao, n.d.).

Id Κλάσης	Ασθένεια	Αριθμός Εμφάνισεων
1	Ηπατίτιδα	2
1	Γρίπη	1
1	Γαστρεντερίτιδα	1
2	Γρίπη	2
2	Γαστρεντερίτιδα	2
2	Ηπατίτιδα	1
2	Βρογχίτιδα	1

Πίνακας 13: Πίνακας με τον αριθμό εμφανίσεων του ευαίσθητου χαρακτηριστικού

Κεφάλαιο 4

Πρακτική εφαρμογή

4.1 Εισαγωγή

Στο παρόν κεφάλαιο, θα εφαρμόσουμε τις τεχνικές ανωνυμοποίησης σε ένα πραγματικό σενάριο, χρησιμοποιώντας μία δημόσια προσβάσιμη λίστα προσωπικών δεδομένων. Συγκεκριμένα, θα παρουσιάσουμε την εφαρμογή τεχνικών ανωνυμοποίησης στο δημόσιο Εθνικό Μητρώο Επιτελικών Στελεχών Δημόσιας Διοίκησης, το οποίο διατίθεται στο διαδικτυακό τόπο του Ανωτάτου Συμβουλίου Επιλογής Προσωπικού (Α.Σ.Ε.Π.) (www.asep.gr). Στο εν λόγω μητρώο εγγράφονται εθελοντικά όσα στελέχη της Δημόσιας Διοίκησης πληρούν συγκεκριμένα κριτήρια βάσει των τυπικών προσόντων τους: τα κριτήρια αυτά προβλέπονται σε ειδική νομοθετική διάταξη (ν. 4369/2016). Το μητρώο αυτό είναι δυναμικά μεταβαλλόμενο, δεδομένου ότι οποιοσδήποτε πληροί τα προσόντα μπορεί να εγγραφεί σε αυτό οποτεδήποτε. Στην παρούσα διατριβή, αξιοποιήσαμε το στιγμιότυπο του εν λόγω μητρώου κατά την 21/10/2016.

Οι λόγοι επιλογής του εν λόγω μητρώου για την παρούσα έρευνα είναι οι εξής:

- α) Είναι ένα μητρώο με επικαιροποιημένα σύγχρονα δεδομένα, ανοιχτής πρόσβασης σε όλους
- β) Όσοι εμφανίζονται στο εν λόγω μητρώο έχουν εγγραφεί με δική τους βούληση, γνωρίζοντας ότι πρόκειται για ένα δημόσια προσβάσιμο μητρώο
- γ) Το εν λόγω μητρώο δεν φέρει κανένα «αρνητικό πρόσημο» για τους εγγεγραμμένους σε αυτό – δηλαδή δεν μπορεί να προκαλέσει την εξαγωγή κάποιου δυσμενούς συμπεράσματος για αυτούς: τουναντίον, η εγγραφή κάποιου σε αυτό υποδηλώνει ότι διαθέτει ένα σύνολο ποιοτικών τυπικών προσόντων.

Στο εν λόγω μητρώο δεν υπάρχει κάποιο γνώρισμα που θα μπορούσε να χαρακτηριστεί ως ευαίσθητο, έτσι ώστε να εφαρμοστεί μία τεχνική ανωνυμοποίησης με γνώμονα την προστασία της ιδιωτικότητας ως προς αυτό το γνώρισμα. Για το λόγο αυτό, και δεδομένου ότι στόχος είναι να εξαχθούν συμπεράσματα ως προς την αποτελεσματικότητα και τους περιορισμούς των τεχνικών ανωνυμοποίησης, θα προσθέσουμε εικονικά ένα νέο γνώρισμα με το χαρακτηρισμό «ευαίσθητο», το οποίο θα λαμβάνει τρεις διαφορετικές τιμές: Α, Β και Γ. Η υπόθεσή μας θα είναι ότι επιθυμούμε να «προστατεύσουμε» τα πρόσωπα της λίστας ως προς την τιμή που έχουν στο συγκεκριμένο ευαίσθητο γνώρισμα (το οποίο βέβαια, δεδομένου ότι είναι εικονικό, δεν αντιπροσωπεύει στην πραγματικότητα απολύτως τίποτα).

Προς αυτήν την κατεύθυνση, θα περιγράψουμε δύο σενάρια:

1. Θα αναθέσουμε τυχαία τιμές (Α, Β ή Γ) σε κάθε εγγραφή του μητρώου στο ευαίσθητο γνώρισμα, με κριτήριο όμως την ισοκατανομή των τιμών αυτών (δηλαδή το 1/3 των εγγραφών θα φέρει την τιμή «Α», το 1/3 την τιμή «Β» και το 1/3 την τιμή «Γ»).
2. Θα αναθέσουμε τυχαία τιμές (Α, Β ή Γ) σε κάθε εγγραφή του μητρώου στο ευαίσθητο γνώρισμα, με κριτήριο όμως την πλήρως ανομοιόμορφη κατανομή των τιμών αυτών (δηλαδή το 60% των εγγραφών θα φέρει την τιμή «Α», το 30% την τιμή «Β» και το 10% την τιμή «Γ»).

Στόχος μας είναι να επιχειρήσουμε τεχνικές ανωνυμοποίησης, προκειμένου να διαπιστωθεί σε τι βαθμό επηρεάζει πραγματικά η κατανομή των τιμών του ευαίσθητου πεδίου την αποτελεσματικότητα της ανωνυμοποίησης.

4.2 Εφαρμογή Τεχνικής Κ-Ανωνυμίας

Για την εφαρμογή της Κ-Ανωνυμίας όπως αναφέρθηκε χρησιμοποιήθηκε η λίστα του δημοσίου Εθνικού Μητρώου Επιτελικών Στελεχών Δημόσιας Διοίκησης η οποία αποτελείται από δεκατέσσερα χαρακτηριστικά τα οποία είναι:

1. Μοναδιαίος Αύξοντας Αριθμός για κάθε εγγραφή
2. Όνομα

3. Επώνυμο
4. Πατρώνυμο
5. Αριθμός ταυτότητας,
6. Προϋποθέσεις εγγραφής,
7. Φορέας οργανικής θέσης,
8. Προϊστάμενος γενικής διεύθυνσης
9. Προϊστάμενος διεύθυνσης,
10. Πτυχίο
11. Μεταπτυχιακό
12. Διδακτορικό
13. Αποφοίτηση ΑΔΙΣΠΟ/ΑΣΠ/ΝΣΠ/ΣΠΑ/Α
14. Ξένη γλώσσα

4.2.1 Διαδικασία Εφαρμογής τεχνικής Κ-Ανωνυμίας

Όπως αναφέρθηκε και παραπάνω ο πίνακας του Εθνικού Μητρώου Επιτελικών Στελεχών Δημόσιας Διοίκησης αποτελείται από δεκατέσσερα πεδία. Για την ερευνά μας επιλέξαμε να στηριχθούμε σε τέσσερα από αυτά τα οποία είναι: {Φορέας Οργανικής Θέσης, Πτυχίο, Μεταπτυχιακό, Διδακτορικό}. Το κριτήριο της επιλογής τους ήταν ο τύπος των τιμών που έπαιρναν αυτά τα χαρακτηριστικά (κατηγορικά) αλλά και η πληρότητά τους σε τιμές. Το γεγονός αυτό θα μας βοηθούσε στο να εξελιχθεί καλύτερα η έρευνά μας και να κατανοηθούν καλύτερα κάποια σημαντικά σημεία και ιδιαιτερότητες που προκύπτουν.

Τα χαρακτηριστικά όπως το {Όνομα, Επώνυμο, Πατρώνυμο, Αριθμός Ταυτότητας} απαλείφθηκαν γιατί προσδιορίζουν άμεσα την ταυτότητα ενός ατόμου. Επίσης η ελάχιστη δυνατή τιμή του k - όπου k το μέγεθος κάθε κλάσης ισοδυναμίας που θα ανακύψει - είναι $k=3$ για να μειώσουμε την πιθανότητα να γίνει ταυτοποίηση σε μία κλάση ισοδυναμίας η οποία θα αποτελείται από τρεις εγγραφές.

Στις επόμενες υποενότητες επεξηγούνται τα βήματα που ακολουθήθηκαν για να πραγματοποιηθούν οι τεχνικές της Γενίκευσης, της Κατάργησης αλλά και η Προσθήκη Θορύβου για την ομαλοποίηση της Γενίκευσης.

4.2.2 Η διαδικασία της Γενίκευσης

Το πρώτο πεδίο του πίνακα όπου έγινε η Γενίκευση ήταν ο Φορέας Οργανικής Θέσης και αφορούσε την δημόσια υπηρεσία που είχε το κάθε άτομο σαν κύριο φορέα απασχόλησης. Οι υπηρεσίες αυτές με βάση το αντικείμενό τους άνηκαν σε κάποιο υπουργείο ή μερικές από αυτές λειτουργούν ως ανεξάρτητες αρχές. Η επιλογή μας για την γενίκευσή τους βασίστηκε στον αντικείμενο που εξυπηρετούσε η κάθε μία υπηρεσία ή στο υπουργείο που ανήκουν. Μερικές υπηρεσίες για παράδειγμα υπάγονται στο υπουργείο Παιδείας, στο υπουργείο Οικονομικών, στο υπουργείο Εσωτερικών κ.λ.π.

Συνεπώς με βάση αυτή την σκέψη δημιουργήθηκαν οι παρακάτω πέντε γενικευμένες τιμές: {Υγεία, Αυτοδιοίκηση, Παιδεία, Οικονομία, Επικρατείας-Εσωτερικών-Πολιτισμός}.

Με τον ίδιο γνώμονα κινηθήκαμε και για την δημιουργία γενικευμένων τιμών στα πεδία: {Πτυχίο, Μεταπτυχιακό και Διδακτορικό}. Υπήρχαν πάρα πολλές διαφορετικές σχολές και τμήματα. Γι' αυτό και ο τρόπος γενίκευσης τους ήταν να αντιστοιχήσουμε κάθε σχολή στο αντίστοιχο πεδίο κατεύθυνσής της (θετικές, επιστήμες, φιλοσοφικές επιστήμες κλπ).

Έτσι προέκυψαν οι παρακάτω γενικευμένες τιμές που ισχύουν και για τα τρία πεδία: {Επιστήμες Υγείας, Θετικές Επιστήμες, Φιλοσοφικές Επιστήμες, Οικονομικών Πολιτικών και Κοινωνικών Επιστημών, Νομικές Επιστήμες}. Αξίζει να αναφερθεί ότι πολλά κελιά σε αυτά τα πεδία δεν είχαν τιμή γι' αυτό και τους δόθηκε η τιμή: {Κενό}.

Αξίζει να σημειωθεί ότι με τη γενίκευση που επιλέξαμε, αναμένεται ότι θα επιτυγχάνουμε μικρή απώλεια πληροφορίας, δεδομένου ότι είναι κατά κανόνα γενίκευση «ενός επιπέδου» με όρους δέντρου ταξινόμησης (με εξαίρεση το γενικευμένο πεδίο «Επικρατείας- Εσωτερικών-Πολιτισμός» στο χαρακτηριστικό «Φορέας»).

4.2.3 Η διαδικασία της Κατάργησης

Η τεχνική της Κατάργησης εφαρμόστηκε από την στιγμή που είχαν σχηματιστεί οι κλάσεις ισοδυναμίας στον πίνακα. Εβδομήντα Οκτώ (78) εγγραφές από τις τετρακόσιες εβδομήντα δύο (470) κρίθηκαν ως ακραίες τιμές οπότε εφαρμόστηκε η τεχνική της Κατάργησης σε επίπεδο εγγραφής και απαλείφθηκαν οριστικά από τον πίνακα. Το νούμερο αυτό αντιστοιχεί σε ένα ποσοστό της τάξης του 16,5%.

Μετά και από την εφαρμογή αυτής της τεχνικής το σύνολο των εγγραφών του πίνακα που ήταν διαθέσιμο για περαιτέρω επεξεργασία ήταν τριακόσιες ενενήντα δύο (392). Συνεπώς το επόμενο βήμα ήταν η προσθήκη θορύβου.

4.2.4 Προσθήκη Θορύβου

Επειδή η τεχνική της Κατάργησης δεν θεωρείται πανάκεια, μερικές εγγραφές επιλέχθηκαν να μην διαγραφούν αλλά να υποστούν μια μικρή τροποποίηση (προσθήκη «θορύβου»). Ο λόγος ήταν επειδή οι συγκεκριμένες εγγραφές, στις τιμές των ψευδοαναγνωριστικών τους προσομοιάζαν σε πολύ μεγάλο ποσοστό στις αντίστοιχες τιμές άλλων εγγραφών και η διαφορά τους ήταν μόνο στην τιμή ενός ψευδοαναγνωριστικού. Για παράδειγμα αν υποθεθεί η εγγραφή Α με τις τιμές: {Φορέας Οργανικής Θέσης: Επικρατείας-Εσωτερικών και Πολιτισμός, Πτυχίο: Φιλοσοφικές Επιστήμες, Μεταπτυχιακό: Φιλοσοφικές Επιστήμες, Διδακτορικό: Φιλοσοφικές Επιστήμες} και η εγγραφή Β με τις τιμές: {Φορέας Οργανικής Θέσης: Επικρατείας-Εσωτερικών και Πολιτισμός, Πτυχίο: Φιλοσοφικές Επιστήμες, Μεταπτυχιακό: Φιλοσοφικές Επιστήμες, Διδακτορικό: Κενό} τότε και η τιμή “Κενό” σε αυτό το κελί θα αντικατασταθεί με την τιμή Φιλοσοφικές Επιστήμες.

Η Προσθήκη Θορύβου χρησιμοποιείται αφενός μεν για να μην διαγραφούν κι άλλες εγγραφές από τον πίνακα και αφετέρου δε για να είναι το δείγμα μας πιο ομοιογενές. Συνολικά στον πίνακά μας προστέθηκαν είκοσι τρία σημεία θορύβου – αριθμός που αντιστοιχεί σε ποσοστό της τάξης του 5,8% των εγγραφών. Κατόπιν αυτού του βήματος προχωρήσαμε στην δημιουργία ενός

εικονικού ευαίσθητου χαρακτηριστικού το οποίο θα παίρνει τρεις διαφορετικές τιμές A, B και Γ. Συνολικά δημιουργήθηκαν δύο διαφορετικά πεδία για το ευαίσθητο χαρακτηριστικό. Το ένα πεδίο περιλαμβάνει την ομοιογενή κατανομή των A, B και Γ, δηλαδή την κατανομή των τιμών σε ίδια ποσοστά εμφάνισης ενώ το άλλο πεδίο περιλαμβάνει την ανομοιογενή κατανομή των τιμών σε διαφορετικά ποσοστά εμφάνισης.

4.2.5 Το ευαίσθητο χαρακτηριστικό

Η συγκεκριμένη λίστα προσωπικών δεδομένων δεν έχει κάποιο ευαίσθητο χαρακτηριστικό. Αυτό σημαίνει ότι μπορεί να υποστεί ανωνυμοποίηση κατά την έννοια της k-ανωνυμοποίησης (δηλαδή να δημιουργηθούν κλάσεις ισοδυναμίας πληθικότητας τουλάχιστον k, έτσι ώστε κάθε εγγραφή να είναι αξεχώριστη από τουλάχιστον άλλες k-1 εγγραφές), αλλά όχι άλλου τύπου ανωνυμοποίηση (π.χ. L-διαφορετικότητας, αφού αυτή η ανωνυμοποίηση αποσκοπεί στο να μην εξάγεται συμπέρασμα περί της τιμής του ευαίσθητου χαρακτηριστικού). Σε συνέχεια λοιπόν των ανωτέρω βημάτων και για τους σκοπούς της παρούσας έρευνας προχωρήσαμε στην δημιουργία ενός εικονικού ευαίσθητου χαρακτηριστικού το οποίο θα παίρνει τρεις διαφορετικές τιμές A, B και Γ. Επιλέξαμε επίσης δύο διαφορετικά σενάρια ως προς το πώς κατανέμονται οι εν λόγω τιμές του νέου αυτού ευαίσθητου χαρακτηριστικού. Στο ένα θεωρήσαμε ομοιογενή κατανομή των A, B και Γ, δηλαδή την κατανομή των τιμών σε ίδια ποσοστά εμφάνισης, ενώ στο δεύτερο σενάριο θεωρήσαμε πλήρως ανομοιογενή κατανομή των τιμών σε διαφορετικά ποσοστά εμφάνισης. Απώτερος στόχος είναι να καταδείξουμε το πώς επηρεάζει η κατανομή των τιμών του ευαίσθητου πεδίου την αποτελεσματικότητα της ανωνυμοποίησης. Με την δημιουργία και των πεδίων του ευαίσθητου χαρακτηριστικού, προχωρήσαμε στην εφαρμογή της τεχνικής της L - Διαφορετικότητας.

4.3 Εφαρμογή L - Διαφορετικότητας

Για να καταδείξουμε σε τι βαθμό επηρεάζει η κατανομή των τιμών του ευαίσθητου χαρακτηριστικού την αποτελεσματικότητα της ανωνυμοποίησης επιλέξαμε την τεχνική της εντροπίας. Η τεχνική της

εντροπίας είναι ένα χρήσιμο εργαλείο που καταδεικνύει τον δείκτη αβεβαιότητας που λαμβάνεται από μία κλάση ισοδυναμίας. Για παράδειγμα όσο περισσότερες ίδιες τιμές βρίσκονται μέσα σε μία κλάση ισοδυναμίας τόσο πιο μικρή θα γίνεται και η εντροπία και κατ' επέκταση και ο δείκτης αβεβαιότητας. Όσο πιο ομοιογενής είναι η κατανομή των τιμών σε μία κλάση ισοδυναμίας τόσο πιο μεγάλη θα είναι και η εντροπία.

Στον πίνακα που βρίσκεται στο Παράρτημα Β και αποτελεί την βάση για την έρευνά μας αυτή, σχηματιστήκαν συνολικά πενήντα επτά κλάσεις ισοδυναμίας για το κάθε ευαίσθητο πεδίο. Όπως έχει ήδη αναφερθεί σε κάθε ευαίσθητο πεδίο οι τιμές είναι τοποθετημένες με διαφορετική κατανομή. Στο πρώτο πεδίο η τιμές είναι τοποθετημένες με ομοιογενή κατανομή (ίδια ποσοστά εμφάνισης των Α, Β και Γ) ενώ στο δεύτερο πεδίο με ανομοιογενή κατανομή (διαφορετικά ποσοστά εμφάνισης).

Στόχος μας είναι να υπολογιστεί η εντροπία σε κάθε κλάση ισοδυναμίας και να εντοπιστούν οι κλάσεις εκείνες οι οποίες παρουσιάζουν χαμηλό βαθμό εντροπίας ενώ ο βαθμός πληροφορίας που ανακτούμε είναι μικρός. Θα πρέπει να αναφερθεί ότι εφόσον το $l=3$, λόγω των τριών χαρακτηριστικών Α, Β και Γ τότε η μέγιστη δυνατή εντροπία που μπορεί να επιτευχθεί είναι $\log(3)$. Στην προκειμένη περίπτωση η μέγιστη δυνατή εντροπία που μπορούμε να επιτύχουμε είναι $\log(3) = 1.584$.

4.3.1 Μετρήσεις στο Πεδίο 1 (Ομοιογενής Κατανομή)

Στο Πεδίο 1 κατόπιν των μετρήσεών μας εντοπίστηκαν συνολικά τέσσερις κλάσεις ισοδυναμίας οι οποίες πέτυχαν τον μέγιστο δυνατό βαθμό της εντροπίας. Χαρακτηριστικό παράδειγμα είναι οι δύο από τις τέσσερις κλάσεις ισοδυναμίας με αύξοντα αριθμό επτά και οκτώ με μόλις τρεις εγγραφές, όπου η κάθε μία περιείχε το χαρακτηριστικό Α, Β και Γ αντίστοιχα.

Επιπλέον ένα άλλο χαρακτηριστικό γνώρισμα των μετρήσεων ήταν ότι το 75% των βαθμών της εντροπίας των εγγραφών πλησίαζε την μέγιστη δυνατή εντροπία ενώ το υπόλοιπο 25% ήταν μικρότερη της μονάδας και κυρίως εμφανιζόταν σε κλάσεις ισοδυναμίες που το πλήθος των εγγραφών

τους δεν ξεπερνούσε τις επτά. Συνεπώς, περίπου το 75% των κλάσεων ισοδυναμίας παρέχει σχεδόν την ιδανική ανωνυμία, αφού ακόμα και αν ο επιτιθέμενος αναγνωρίσει ότι ένα συγκεκριμένο πρόσωπο βρίσκεται σε μία εξ αυτών των κλάσεων ισοδυναμίας, δεν θα μπορέσει να εξάγει κανένα συμπέρασμα αναφορικά με την τιμή του ευαίσθητου πεδίου.

Στον πίνακα 14 απεικονίζονται μερικές από τις κλάσεις από τις κλάσεις με την μεγαλύτερη εντροπία.

	ID κλάσης	Βαθμός Εντροπίας
1	Κλάση 7	1,584
2	Κλάση 8	1,584
3	Κλάση 31	1,584
4	Κλάση 47	1,584
5	Κλάση 44	1,579
6	Κλάση 14	1,579
7	Κλάση 40	1,570
8	Κλάση 22	1,562
9	Κλάση 23	1,562
10	Κλάση 50	1,562

Πίνακας 14: Οι κλάσεις με τους μεγαλύτερους δείκτες εντροπίας

4.3.2 Μετρήσεις στο Πεδίο 2 (Ανομοιογενής Κατανομή)

Στο Πεδίο 2 είχαμε την ανομοιογενή κατανομή των χαρακτηριστικών γι' αυτό και σε αρκετές περιπτώσεις μερικές κλάσεις ισοδυναμίας περιείχαν μόνο το "Α" και το "Β" (το οποίο καθιστά τις εγγραφές σε αυτές τις κλάσεις ισοδυναμίας ευάλωτες σε επιθέσεις εξαγωγής συμπεράσματος, αφού αν προσδιορίσουμε ότι κάποιο πρόσωπο ανήκει σε μία από αυτές τις κλάσεις συμπεραίνουμε αυτόματα ότι αποκλείεται να έχει την τιμή «Γ» στο ευαίσθητο πεδίο). Παρατηρήσαμε επίσης ότι λόγω αυτού του γεγονότος το 60% εγγραφών είχαν τον δείκτη της εντροπίας μικρότερο από την μονάδα ενώ το υπόλοιπο 40% ήταν μεγαλύτερο ή ίσο της μονάδας.

Επιπλέον η κλάση με τον μικρότερο δείκτη εντροπίας ήταν η κλάση είκοσι επτά της οποίας η εντροπία ήταν ίση με το μηδέν. Αυτό συμβαίνει φυσικά γιατί στην κλάση αυτή κυριαρχούσε ένα χαρακτηριστικό και αυτό ήταν το “Α”. Ουσιαστικά, στην περίπτωση αυτή, για τις εγγραφές αυτής της κλάσης ισοδυναμίας δεν επιτυγχάνεται καμία ανωνυμοποίηση – αφού, άπαξ και προσδιοριστεί ότι ένα πρόσωπο εντάσσεται σε αυτήν την κλάση γνωρίζουμε αμέσως, με βεβαιότητα, την τιμή στο ευαίσθητο χαρακτηριστικό του. Παρ’ όλη την ιδιομορφία του Πεδίου 2 - και το γεγονός της ανομοιογένειας - δημιουργήθηκαν δύο κλάσεις ισοδυναμίας οι οποίες είχαν την μέγιστη δυνατή εντροπία (1.584). Είναι οι κλάσεις δέκα και τριάντα έξι οι οποίες αποτελούνται από τρεις εγγραφές που περιέχουν και τα τρία χαρακτηριστικά.

Στον πίνακα 15 απεικονίζονται οι κλάσεις με τους μικρότερους δείκτες εντροπίας.

	ID κλάσης	Βαθμός Εντροπίας
1	Κλάση 27	0
2	Κλάση 32	0,720
3	Κλάση 42	0,720
4	Κλάση 22	0,812
5	Κλάση 28	0,812
6	Κλάση 37	0,812
7	Κλάση 39	0,812
8	Κλάση 46	0,812
9	Κλάση 52	0,812
10	Κλάση 15	0,915

Πίνακας 15: Οι κλάσεις με τους μικρότερους δείκτες εντροπίας

Κεφάλαιο 5

Επίλογος

Η συγκεκριμένη μεταπτυχιακή διατριβή ασχολήθηκε με το πρόβλημα της διατήρησης της ιδιωτικότητας σε μεγάλα σύνολα δεδομένων. Παρουσίασε την έννοια της ανωνυμοποίησης και ανέλυσε μερικές από τις βασικότερες τεχνικές που έχουν αναπτυχθεί μέχρι σήμερα για την εφαρμογή της.

Στην συνέχεια εφήρμοσε δύο από τις τεχνικές ανωνυμοποίησης, την K-Ανωνυμία και L - Διαφορετικότητα σε ένα δημόσια προσβάσιμο πίνακα του ΑΣΕΠ. Τα συμπεράσματα ήταν ιδιαίτερα χρήσιμα και καταδεικνύουν με τον πιο σαφή τρόπο πως μπορεί η κατανομή του ευαίσθητου χαρακτηριστικού σε έναν πίνακα να επηρεάσει τον βαθμό επιτυχίας της ανωνυμοποίησης.

Τέλος είδαμε πως με την ομοιογενή κατανομή του ευαίσθητου χαρακτηριστικού κάποιος επιτιθέμενος είναι πιο δύσκολο να εξάγει συμπεράσματα για την ταυτότητα ενός ατόμου σε σχέση με την ανομοιογενή κατανομή εφόσον τα ποσοστά εμφάνισης ενός χαρακτηριστικού είναι πιο μεγάλα.

Ως εκ τούτου, δεν υπάρχει κάποια τεχνική ανωνυμοποίησης που να διασφαλίζει εγγυημένα ένα βέλτιστο αποτέλεσμα από την πλευρά της ανωνυμοποίησης, αλλά θα πρέπει κάθε φορά να διερευνάται το περιεχόμενο του πίνακα δεδομένων που πρόκειται να ανωνυμοποιηθεί, προκειμένου να προσαρμόζεται κατάλληλα η τεχνική ανωνυμοποίησης που θα επιλεγεί.

Παράρτημα Α

Πίνακες

1. **Πίνακας 1:** Γενικευμένο Αριθμητικό Πεδίο
2. **Πίνακας 2:** Νοσοκομειακή Λίστα Ασθενών
3. **Πίνακας 3:** Ανωνυμοποιημένος Πίνακας με $k=2$
4. **Πίνακας 4:** Πίνακας που έχει εφαρμοστεί η Κατάργηση
5. **Πίνακας 5:** Παράδειγμα k - Ανωνυμίας με $k=3$
6. **Πίνακας 6:** Ανωνυμοποιημένος Πίνακας με $k=2$
7. **Πίνακας 7:** Πίνακας με L-Diversity 2
8. **Πίνακας 8:** Πίνακας 16: Στοιχεία Μισθού και Ασθένεια
9. **Πίνακας 9:** Ανωνυμοποιημένος Πίνακας με L-Diversity=3
10. **Πίνακας 10:** Νοσοκομειακή Λίστα Ασθενών
11. **Πίνακας 11:** Πίνακας με k -Ανωνυμία=4 και L-Διαφορετικότητα=3
12. **Πίνακας 12:** Πίνακας με το ID της κλάσης ισοδυναμίας
13. **Πίνακας 13:** Πίνακας με τον αριθμό εμφανίσεων του ευαίσθητου χαρακτηριστικού
14. **Πίνακας 14:** Οι κλάσεις με τους μεγαλύτερους δείκτες εντροπίας
15. **Πίνακας 15:** Οι κλάσεις με τους μικρότερους δείκτες εντροπίας

Παράρτημα Β

Ανωνυμοποιημένος Πίνακας

ID	Φορέας Οργανικής Θέσης	Πτυχίο	Μεταπτυχιακό	Διδακτορικό	Πεδίο1 (Ομοιογενές)	Πεδίο 2 (Μη ομοιογενές)	ID Κλάσης
1	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	A	B	1
2	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	B	A	
3	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	B	B	
4	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	A	Γ	
5	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	Γ	A	
6	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	B	B	
7	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	A	A	2
8	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	A	
9	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	B	
10	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	A	B	
11	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	B	
12	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	A	
13	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	A	B	
14	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	Γ	
15	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	A	
16	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	B	
17	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	A	A	
18	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	B	
19	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	B	
20	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	A	
21	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	B	A	
22	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	A	A	

23	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Β	3
24	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Α	
25	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	Α	Α	
26	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	Β	Β	
27	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	Β	Α	
28	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	Γ	Γ	4
29	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	Α	Α	
30	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Β	
31	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Α	Β	
32	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Α	Α	
33	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Α	5
34	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Α	
35	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Β	
36	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Α	Α	
37	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Β	
38	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Α	6
39	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Α	
40	ΥΓΕΙΑ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Γ	
41	ΥΓΕΙΑ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Β	Β	7
42	ΥΓΕΙΑ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Β	
43	ΥΓΕΙΑ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Α	Α	
44	ΥΓΕΙΑ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Α	8
45	ΥΓΕΙΑ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Α	
46	ΥΓΕΙΑ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Β	
47	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Α	9

		ΚΟΙΝ.					
48	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Β	Α	
49	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Β	
50	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Α	Α	
51	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Β	
52	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Β	Α	
53	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Α	
54	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΗ ΥΓΕΙΑΣ	ΚΕΝΟ	Β	Β	
55	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	Α	Γ	10
56	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	Γ	Β	
57	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	Α	Α	
58	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Α	11
59	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Γ	
60	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Β	
61	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Α	
62	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Α	
63	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Β	
64	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Β	
65	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Α	12
66	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Α	
67	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ.	ΚΕΝΟ	ΚΕΝΟ	Α	Β	

		ΚΟΙΝ.					
68	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Γ	Β	
69	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Β	Γ	
70	ΥΓΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Β	Α	
71	ΥΓΕΙΑ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Γ	Β	
72	ΥΓΕΙΑ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Α	Α	
73	ΥΓΕΙΑ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Β	Α	
74	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Β	13
75	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Β	Β	
76	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Α	Α	
77	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Β	Α	
78	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Α	
79	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Β	
80	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Α	Α	
81	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Β	Γ	
82	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Α	Α	
83	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΚΕΝΟ	Γ	Α	
84	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Β	14
85	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Γ	Α	
86	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Γ	Β	
87	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Α	
88	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Α	Α	
89	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Γ	Γ	
90	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Γ	Β	
91	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Α	
92	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Β	
93	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Γ	Α	
94	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Α	Α	
95	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Γ	
96	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ	ΘΕΤΙΚΕΣ	ΘΕΤΙΚΕΣ	Α	Α	

127	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Β	
128	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Β	
129	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Α	Α	
130	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Α	
131	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Β	
132	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Α	Α	
133	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Β	
134	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Α	Α	
135	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Α	
136	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Α	17
137	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Β	
138	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Α	
139	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Γ	18
140	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Α	
141	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Α	Α	
142	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Α	
143	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Γ	Β	
144	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Α	Α	
145	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Β	Α	
146	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Γ	Α	
147	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Α	Α	19
148	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Β	Β	
149	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Α	Α	
150	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Β	Β	
151	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Γ	Α	
152	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Β	Α	
153	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΘΕΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Α	Β	
154	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Α	20
155	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΦΙΛΟΣΟΦΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Α	Α	

		ΚΟΙΝ.	ΚΟΙΝ.				
200	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Β	
201	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Α	
202	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Α	
203	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Β	
204	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Α	
205	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Γ	Α	
206	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Α	Α	
207	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	Β	Β	
208	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Γ	Α	24
209	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Α	Γ	
210	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Γ	Β	
211	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Β	Α	
212	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Α	Α	
213	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Β	25
214	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Β	
215	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Γ	Α	
216	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Α	Α	26
217	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Α	Γ	
218	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Γ	Α	
219	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	ΚΕΝΟ	Β	Β	
220	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΚΕΝΟ	ΚΕΝΟ	ΚΕΝΟ	Α	Α	27
221	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΚΕΝΟ	ΚΕΝΟ	ΚΕΝΟ	Α	Α	
222	ΑΥΤΟΔΙΟΙΚΗΣΗ	ΚΕΝΟ	ΚΕΝΟ	ΚΕΝΟ	Β	Α	
223	ΠΑΙΔΕΙΑ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	ΕΠΙΣΤΗΜΕΣ ΥΓΕΙΑΣ	Γ	Α	28

	ΠΟΛΙΤΙΣΜΟΣ	ΚΟΙΝ.	ΚΟΙΝ.				
387	ΕΠΙΚΡΑΤΕΙΑΣ ΕΩΤΕΡΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΣΜΟΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Β	Β	56
388	ΕΠΙΚΡΑΤΕΙΑΣ ΕΩΤΕΡΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΣΜΟΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Β	Α	
389	ΕΠΙΚΡΑΤΕΙΑΣ ΕΩΤΕΡΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΣΜΟΣ	ΕΠΙΣΤΗΜΕΣ ΟΙΚΟΝ. ΠΟΛΙΤ. ΚΟΙΝ.	ΚΕΝΟ	ΚΕΝΟ	Γ	Α	
390	ΕΠΙΚΡΑΤΕΙΑΣ ΕΩΤΕΡΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΣΜΟΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Α	57
391	ΕΠΙΚΡΑΤΕΙΑΣ ΕΩΤΕΡΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΣΜΟΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Α	Β	
392	ΕΠΙΚΡΑΤΕΙΑΣ ΕΩΤΕΡΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΣΜΟΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΝΟΜΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΚΕΝΟ	Β	Β	

Βιβλιογραφία

1. Alexin, Z. (2014). Does fair anonymization exist? *International Review of Law, Computers & Technology*, 28(1), 21–44.
<http://doi.org/10.1080/13600869.2013.869909>
2. Li, N., Li, T., & Venkatasubramanian, S. (2007). t -Closeness : Privacy Beyond k -Anonymity and l -Diversity. *IEEE 23rd International Conference on Data Engineering*, (3), 106–115. <http://doi.org/10.1109/ICDE.2007.367856>
3. M Fung, B. C., Wang, K., Chen, R., Yu, P. S., & M Fung, B. C. (2010). Privacy-Preserving Data Publishing: A Survey of Recent Developments ACM Reference Format. *ACM Computing Surveys P. S. ACM Comput. Surv*, 42(53).
<http://doi.org/10.1145/1749603.1749605>
4. Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset).
5. Samarati, P. (2001a). Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*.
<http://doi.org/10.1109/69.971193>
6. Singh, A. P., Philosophy, D. O. F., & Science, C. (2012). An Iterative Approach to Examining the Effectiveness of Data Sanitization.
7. Sweeney, L. (2002a). Achieving k -anonymity Privacy Protection using Generalization and Suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 571–588.
<http://doi.org/10.1142/S021848850200165X>
8. Sweeney, L. (2002b). k -ANONYMITY: A MODEL FOR PROTECTING PRIVACY. *International Journal on Uncertainty*, 10(5), 557–570.
9. Xiao, X., & Tao, Y. (n.d.). Anatomy: Simple and Effective Privacy Preservation.