Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλ. Μηχανικών & Μηχανικών Υπολογιστών
Βάσεις Δεδομένων
Εαρινό Εξάμηνο 2021-2022

Ομάδα Project 109
Ιωάννης Δρέσσος - 03119608
Ιωάννης Μποσκοβιτς - 03119640
Γεώργιος Κίτσιος - 03119801

Αναφορά Εξαμηνιαίας Εργασίας

Εισαγωγή

Η εργασία αφορά την ανάπτυξη εφαρμογής γραφικής διεπαφής χρήστη (GUI application) για σύνδεση με βάση δεδομένων. Ο χρήστης μέσω της εφαρμογής θα μπορεί να περιηγηθεί, να προβάλει και να επεξεργαστεί τα δεδομένα που περιέχονται στη βάση, χωρίς να απαιτείται εκ μέρους του κάποια τεχνική γνώση.

Επιλέξαμε η εφαρμογή να μην είναι διαδικτυακή, αλλά να είναι εγγενής ώστε η εγκατάσταση και παραμετροποίηση της απο το μηδέν να είναι εύκολη και με λίγα προαπαιτούμενα, ακόμη και για κάποιον που δεν έχει εμπειρία με σχετικές διαδικασίες.

Τεχνικές Προδιαγραφές

Αποθετήριο GitHub: https://github.com/idressos/hfri-erp

Η εφαρμογή έχει αναπτυχθεί στην γλώσσα προγραμματισμού Java. Η επιλογή αυτή της γλώσσας προσφέρει τα παρακάτω πλεονεκτήματα:

- Εύκολη διαδικασία μεταγλώττισης πηγαίου κώδικα - επιτάχυνση της φάσης ελέγχου.
- Το αρχείο που προκύπτει απο την μεταγλώττιση του πηγαίου κώδικα της εφαρμογής είναι εκτελέσιμο σε κάθε λειτουργικό σύστημα που υποστηρίζει το JVM. Δεν απαιτείται η μεταγλώττιση του ξεχωριστά για διαφορετικά λειτουργικά συστήματα.
- Ο πηγαίος κώδικας είναι οργανωμένος και ευανάγνωστος λόγω της αντικειμενοστρεφούς φύσης της Java.
- Οι μέθοδοι/εντολές και γενικότερα το συντακτικό της γλώσσας είναι εύκολα κατανοήσιμα απο οποιονδήποτε με εμπειρία σε άλλες γλώσσες.
- Η γλώσσα έχει άριστη ενσωματωμένη υποστήριξη για SQL.

Για την ανάπτυξη της εφαρμογής χρησιμοποιήθηκαν βιβλιοθήκες που παρέχονται από τον κατασκευαστή η από τρίτους για χρήση σε λογισμικό ανοιχτού κώδικα υπό όρους (licenses), οι οποίοι τηρούνται σε αυτή την ακαδημαϊκή εργασία:

- https://github.com/mariadb-corporation/mariadb-connector-j
    - Χρησιμοποιείται για την σύνδεση και επικοινωνία της εφαρμογής με την βάση δεδομένων.
- https://github.com/stleary/JSON-java
    - Χρησιμοποιείται για την κωδικοποίηση και αποκωδικοποίηση του αρχείου παραμέτρων της εφαρμογής.
- https://github.com/mwiede/jsch
    - Χρησιμοποιείται (προεραιτικά) για την ανακατεύθυνση της σύνδεσης της εφαρμογής με την βάση δεδομένων μέσω σήραγγας SSH (SSH Tunnel).

Οι βιβλιοθήκες αυτές ενσωματώνονται αυτόματα στο εκτελέσιμο κατα τη διαδικασία της μεταγλώττισης και δεν απαιτείται κάποια περαιτέρω ενέργεια από τον χρήστη για την απόκτηση τους.

Η έκδοση της Java που χρησιμοποιήθηκε για την ανάπτυξη της εφαρμογής είναι η 17. Συγκεκριμένα, χρησιμοποιήθηκε το Java SE Development Kit 17 το οποίο είναι η επίσημη υλοποίηση των στάνταρ της Java από τον κατασκευαστή της Oracle κατα τον χρόνο σύνταξης της παρούσας αναφοράς.

Επομένως, για την εκτέλεση της εφαρμογής από τον χρήστη απαιτείται να υπάρχει στο σύστημα εγκατεστημένη μια υλοποίηση του JVM έκδοσης τουλάχιστον 17. Νεότερες εκδόσεις θα υποστηρίζονται σύμφωνα με τις προδιαγραφές του κατασκευαστή του ανάλογου JVM, ενώ παλαιότερες δεν θα υποστηρίζονται.

Οδηγίες Εγκατάστασης

Προαπαιτούμενα:

- Διακομιστής SQL βάσης δεδομένων.
    - **Υποστηρίζονται MySQL 8.0+ και MariaDB 10.0+**
    - Αρχικοποίηση της βάσης της εργασίας με τα DDL/DML scripts

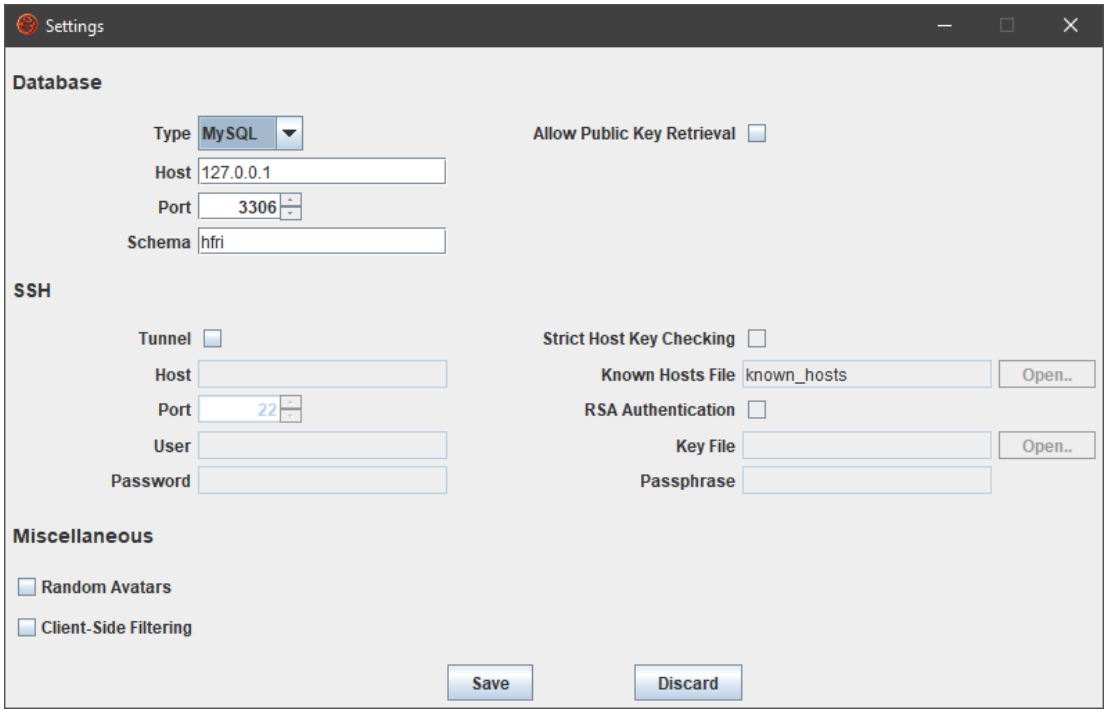
1. Εγκατάσταση οποιασδήποτε υλοποίησης του JVM έκδοσης 17 και άνω.
    1.1. Προτείνεται το επίσημο Java SE Development Kit της Oracle: https://www.oracle.com/java/technologies/downloads/
2. Εγκατάσταση του εργαλείου αυτοματοποίησης κατασκευής Apache Maven: https://maven.apache.org/download.cgi
    2.1. Οδηγίες εγκατάστασης: https://maven.apache.org/install.html

3. Εγκατάσταση του Git: https://git-scm.com/book/en/v2/Getting-Started-Installing-Git
4. Κλωνοποιήση του αποθετηρίου GitHub της εφαρμογής και μεταγλώττιση:

```
$ git clone https://github.com/idressos/hfri-erp.git
$ cd hfri-erp
$ mvn clean install
```

Στην περίπτωση επιτυχημένης μεταγλώττισης, το εκτελέσιμο θα δημιουργηθεί στον φάκελο `target` και θα έχει όνομα της μορφής `hfri-erp-*.jar`. Το αρχείο είναι έτοιμο προς χρήση, και αν επιλέξει ο χρήστης μπορεί να το μεταφέρει σε ξεχωριστό φάκελο και να διαγράψει τους φακέλους και τα αρχεία που προέκυψαν από τις διαδικασίες κλωνοποίησης και μεταγλώττισης.

Παραμετροποίηση Εφαρμογής



Οι ρυθμίσεις της εφαρμογής χωρίζονται σε δύο κατηγορίες:

<u>Ρυθμίσεις Σύνδεσης με Βάση Δεδομένων</u>

Type: Ο τύπος του διακομιστή της βάσης δεδομένων
Host: Η διαδικτυακή διεύθυνση του διακομιστή
Port: Η θύρα στην οποία είναι δεσμευμένος ο διακομιστής
Schema: Το όνομα του διαγράμματος/βάσης όπου θα συνδεθεί η εφαρμογή
Allow Public Key Retrieval: Δείτε παρακάτω

Ρυθμίσεις SSH

Όπως προαναφέρθηκε στις τεχνικές προδιαγραφές της εφαρμογής, ο χρήστης έχει την επιλογή να ανακατευθύνει την σύνδεση της εφαρμογής μέσω σήραγγας SSH. Η επιλογή αυτή μπορεί να εξυπηρετεί τους εξής σκοπούς:

- Κρυπτογράφηση των δεδομένων που μεταφέρονται μέσω της σύνδεσης, σε περίπτωση που ο διακομιστής της βάσης δεδομένων βρίσκεται σε απομακρυσμένο υπολογιστή και δεν είναι τοπικά εγκατεστημένος.
- Σύνδεση σε διακομιστή βάσης δεδομένων που είναι δεσμευμένος τοπικά σε απομακρυσμένο υπολογιστή, και δεν έχει δημόσια IP διεύθυνση (π.χ. μόνο `localhost:3306`).
- Σύνδεση σε διακομιστή βάσης δεδομένων που τρέχει πίσω από τείχος προστασίας (firewall) σε απομακρυσμένο υπολογιστή.

Στην περίπτωση που ο χρήστης επιλέξει να χρησιμοποιήσει SSH tunneling, οι ρυθμίσεις `Database/Host` και `Database/Port` που αναφέρθηκαν παραπάνω προφανώς **πρέπει να προσαρμοστούν στην νεα δικτυακή διαμόρφωση της εφαρμογής**.

→ Για παράδειγμα, αν η διεύθυνση στο πεδίο `Database/Host` είναι 127.0.0.1 (η localhost), δεν δείχνει πια στον τοπικό υπολογιστή του χρήστη που τρέχει την εφαρμογή, αλλά στον διακομιστή SSH.

Σήρραγα SSH μεταξύ 149.32.233.163 και 85.53.130.12, επιτρέπει την επικοινωνία του προσωπικού υπολογιστή με βάση δεδομένων που τρέχει τοπικά στον διακομιστή.

Στο παραπάνω διάγραμμα, σωστή διαμόρφωση είναι η εξής:

```
Database/Host: 127.0.0.1
Database/Port: 3306
SSH/SSH Tunnel: Enabled
SSH/Host: 85.53.130.12
SSH/Port: (Σύμφωνα με τη διαμόρφωση του διακομιστή SSH. Συνήθως 22)
```

Προχωρημένες Ρυθμίσεις

Database/Allow Public Key Retrieval
Επιτρέπει στην εφαρμογή να ζητήσει απευθείας το RSA public key του διακομιστή.

Προσοχή! Δεν συνιστάται διότι κάνει τον κωδικό του χρήστη ευάλωτο σε τυχόν επιθέσεις man-in-the-middle κατά τη σύνδεση.
Προορίζεται για χρήση μόνο σε περίπτωση που προκύπτει σχετικό σφάλμα κατα τη σύνδεση με τον διακομιστή βάσης δεδομένων, και μονο σε τοπική εγκατάσταση για δοκιμές.

SSH/Strict Host Key Checking
Κάθε φορά που συνδεόμαστε σε έναν διακομιστή SSH, μας στέλνει ένα «κλειδί» που λειτουργεί ως αναγνωριστικό για τον συγκεκριμένο διακομιστή, ώστε να γνωρίζουμε ότι όντως έχουμε συνδεθεί στον υπολογιστή που θέλουμε και όχι στον διακομιστή ενός κακόβουλου τρίτου που μας επιτίθεται με σκοπό την υποκλοπή των δεδομένων που θα μεταφέρουμε μέσω της σύνδεσης.

Καλό είναι επομένως, κάποια φορά που θα συνδεθούμε σε έναν διακομιστή SSH μέσω μιας σύνδεσης που εμπιστευόμαστε και θεωρούμε ασφαλή (συνήθως την πρώτη φορά), να αποθηκεύσουμε το κλειδί του ώστε να το συγκρίνουμε με αυτό που μας θα στέλνει κάθε φορά που θα συνδεόμαστε σε αυτόν μελλοντικά.

Έτσι, σε περίπτωση επίθεσης θα ενημερωθούμε οτι η διαδικασία σύγκρισης των κλειδιών απέτυχε, δηλαδή ο κακόβουλος διακομιστής μας έστειλε λάθος αναγνωριστικό κλειδί.

Αν απενεργοποιήσουμε τη ρύθμιση αυτή, η διαδικασία σύγκρισης κλειδιών δεν εκτελείται.
Η ρύθμιση αυτή προορίζεται για χρήση την πρώτη φορά που πρόκειται να συνδεθούμε σε διακομιστή βάσης δεδομένων μέσω σήραγγας SSH, ωστε να μην αποτύχει η σύνδεση και να μας δοθεί η ευκαιρία να αποθηκεύσουμε το κλειδί του διακομιστή SSH στο τοπικό αρχείο που ορίζεται από την ρύθμιση `SSH/Known Hosts File`.

SSH/RSA Authentication

Σε όσους χρησιμοποιούν συχνά SSH για απομακρυσμένη σύνδεση, η παραμετροποιούν οι ίδιοι τους διακομιστές SSH για αυξημένη ασφάλεια, θα είναι γνωστό ότι το πρωτόκολλο SSH επιτρέπει μια εναλλακτκή, πιο ασφαλή μέθοδο ταυτοποίησης χρηστών από κλασικούς κωδικούς (password authentication).
Η μέθοδος αυτή ονομάζεται RSA authentication.

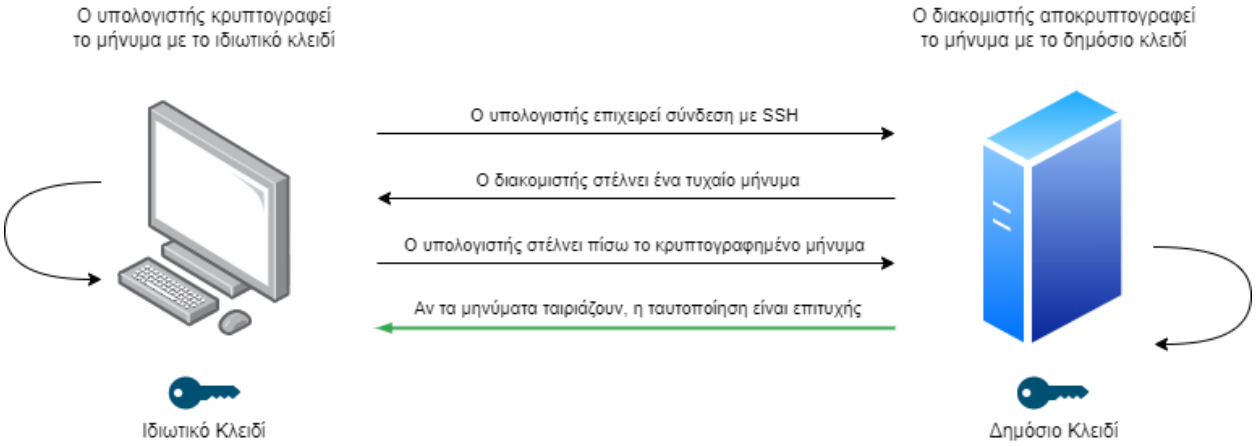
Για να χρησιμοποιήσει κάποιος RSA authentication πρέπει με συγκεκριμένους ασφαλείς αλγορίθμους, να δημιουργήσει ένα ζεύγος κλειδιών, τα οποία ουσιαστικά είναι τυχαίες συμβολοσειρές προδιαγεγραμμένου μεγέθους.

Το ένα κλειδί είναι το ιδιωτικό κλειδί του χρήστη (private key), το οποίο κρατάει μυστικό και δεν το μοιράζεται με κανέναν. Με το κλειδί αυτό «υπογράφει» ο χρήστης μηνύματα που επιθυμεί να στείλει.

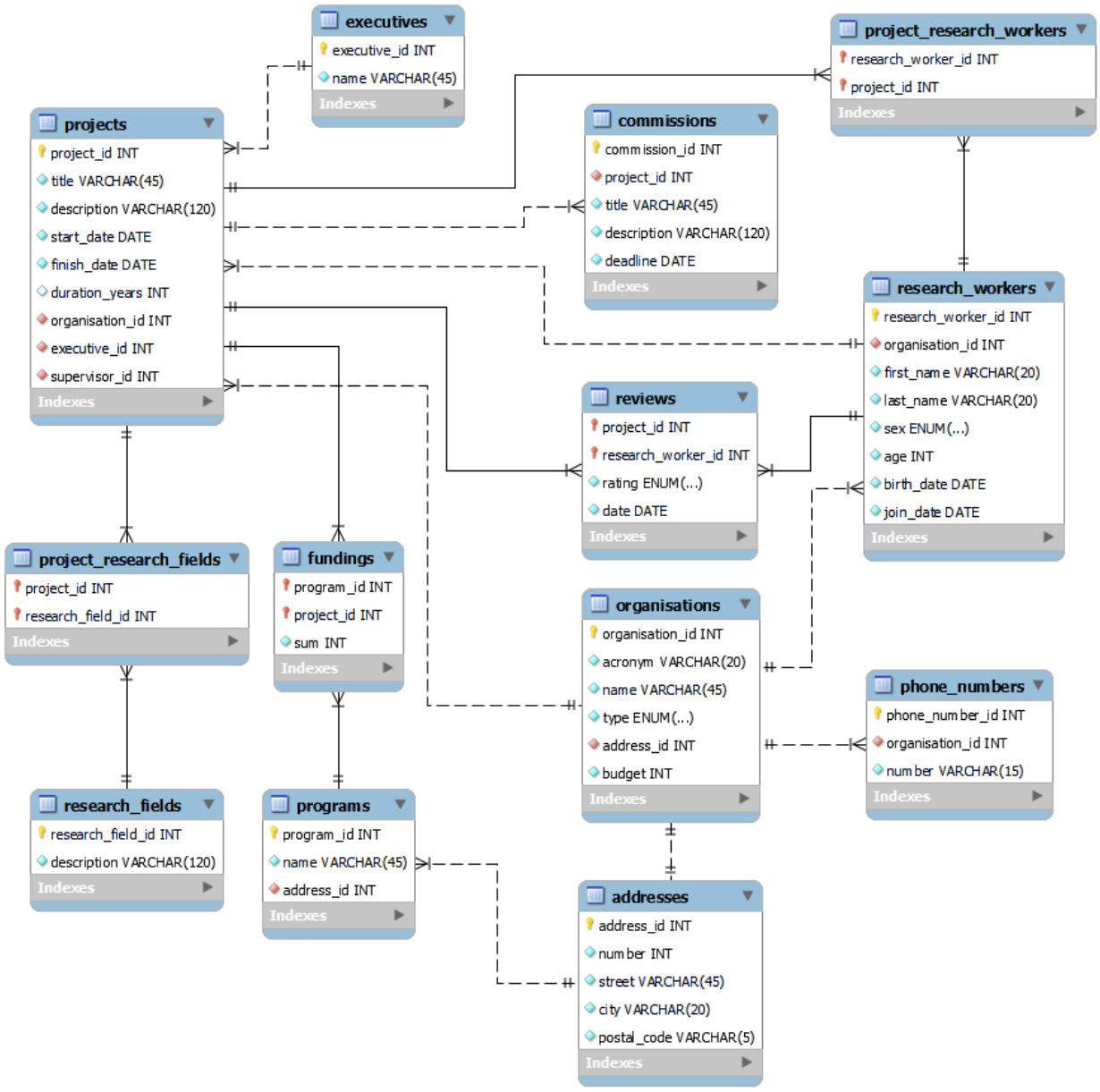
Το άλλο κλειδί είναι το δημόσιο κλειδί του χρήστη (public key), το οποίο χρησιμοποιείται για να επιβεβαιώσει κάποιος ότι ένα μήνυμα το έχει όντως υπογράψει ο κάτοχος του ιδιωτικού κλειδιου, και όχι κάποιος άλλος.

Στην περίπτωση του πρωτοκόλλου SSH, ο χρήστης κατα τη διαδικασία ρύθμισης του διακομιστή SSH μπορεί να ανεβάσει το δημόσιο κλειδί του στον διακομιστή και να ενεργοποιήσει την ταυτοποίηση με κλειδιά RSA.

Η διαδικασία ταυτοποίησης περιγράφεται αναλυτικά από το ακόλουθο διάγραμμα:



Σχεσιακό Διάγραμμα Βάσης Δεδομένων



Το διάγραμμα μας διαφέρει απο το προτεινόμενο της εργασίας με τους εξής τρόπους:

- Τα έργα (projects) διαφέρουν από τις επιχορηγήσεις (fundings). Ένα έργο μπορεί να λάβει μια μόνο επιχορήγηση ακέραιου ποσού από κάθε πρόγραμμα.
    - Πολλά προγράμματα μπορούν να επιχορηγήσουν ένα έργο, εφόσον αυτό έχει αξιολογηθεί.

- Ένας ερευνητής (research worker) μπορεί να υποβάλλει μόνο μία αξιολόγηση (review) για κάθε έργο.
- Τα έργα δεν έχουν πεδίο που να περιγράφει αν είναι ενεργά ή όχι.
    - Αυτό οφείλεται στο οτι σε MySQL διακομιστές δεν επιτρέπεται να χρησιμοποιηθούν μη-ντετερμινιστικές συναρτήσεις όπως `CURDATE()` σε εκφράσεις παραγόμενων πεδίων. Επομένως, δεν υπάρχει τρόπος να υπολογίσουμε βάση της ημερομηνίας λήξης του έργου το αν είναι ενεργό ή όχι.
    - Υπάρχει ωστόσο συνάρτηση για να ελέγξουμε αν ένα έργο είναι ενεργό ή όχι στο πρόγραμμα μας.
- Για τον ίδιο λόγο το πεδίο της ηλικίας στους ερευνητές δεν είναι παραγόμενο, αλλά εισαγόμενο.
- Το πεδίο `supervisor_id` στα έργα είναι foreign key του πεδίου `research_worker_id` απο την οντότητα των ερευνητών.

Σημειώσεις

1. Το ερώτημα 3.8 υλοποιήθηκε για 2 ή περισσότερα έργα που δεν έχουν παραδοτέα αντί για 5 λόγω των ελλιπών δεδομένων απο την τυχαία γεννήτρια.
2. Η ρύθμιση `Miscellaneous/Client-Side Filtering` επιτρέπει όπου είναι δυνατό το φιλτράρισμα των αποτελεσμάτων τοπικά χωρίς αυτά να επικαιροποιηθούν απο την βάση δεδομένων πρώτα.
    2.1. Απο προεπιλογή η ρύθμιση αυτή είναι απενεργοποιημένη διότι στην εκφώνηση ζητείται η επίλυση των ερωτημάτων σε SQL. Οταν η ρύθμιση είναι απενεργοποιημένη, το φιλτράρισμα γίνεται με την χρήση της πρότασης `WHERE`.
    2.2. Η ρύθμιση αυτή χρησιμεύει στο να μειωθούν οι συνδέσεις στη βάση δεδομένων σε περιβάλλον δοκιμής όπου τα δεδομένα δεν μεταβάλλονται δυναμικά.
3. Η ρύθμιση `Miscellaneous/Random Avatars` χρησιμοποιεί το randomuser.me API για να εμφανίσει τυχαίες φωτογραφίες για τους ερευνητικούς υπαλλήλους σε κάθε ανανέωση. Δεν συνιστάται η ενεργοποίηση της σε συνδέσεις χαμηλής ταχύτητας.