A black background with a black square

Description automatically generated with medium confidence**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2023-2024

ΕΞΑΜΗΝΟ 6ο

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

ΥΠΕΥΘΥΝΟΣ ΚΑΘΗΓΗΤΗΣ: Δ. ΤΣΟΥΜΑΚΟΣ

**Βάσεις Δεδομένων**

**Εξαμηνιαία Εργασία**

|  |  |
| --- | --- |
| **ΣΥΓΓΡΑΦΕΙΣ** | Κάκος Σωτήριος  Αριθμός Μητρώου : 03121110  Email: [el21110@mail.ntua.gr](mailto:el21110@mail.ntua.gr) |
| **ΠΕΡΙΕΧΟΜΕΝΑ** | 1. Διάγραμμα Οντοτήτων - Συσχετίσεων 2. Σχεσιακό Σχήμα 3. Ευρετήρια 4. DDL και DML 5. Authentication |

1. **Διάγραμμα Οντοτήτων - Συσχετίσεων**

Ζητούμενο της εργασίας είναι η σχεδίαση και υλοποίηση κατάλληλης βάσης δεδομένων, η οποία θα αξιοποιηθεί από δημοφιλή διαγωνισμό μαγειρικής για την αποθήκευση και διαχείριση σχετικών πληροφοριών.

Η πρώτη φάση της ανάπτυξης της βάσης ήταν η εξόρυξη των απαιτήσεων και των σχετικών προδιαγραφών που θα έπρεπε αυτή να πληροί. Ο εννοιολογικός σχεδιασμός της βάσης στηρίχθηκε γύρω από τις κύριες κατηγορίες πληροφοριών που θέλαμε να αποθηκεύουμε, δηλαδή αυτές σχετικά με τις συνταγές, τους μάγειρες και τα επεισόδιά του διαγωνισμού. Προσπαθώντας να κάνουμε μία αφαιρετική αλλά πλήρη περιγραφή του μικρόκοσμού μας καταλήξαμε διαισθητικά στο διάγραμμα οντοτήτων συσχετίσεων που παρατίθεται παρακάτω.

A diagram of a computer program

Description automatically generated

Ο διαγωνισμός διαδραματίζεται σε επεισόδια σε καθένα από τα οποία επιλέγονται με τυχαίο τρόπο 10 εθνικές κουζίνες, 3 κριτές και 10 συνταγές. Στην υλοποίησή μας χρησιμοποιούμε δύο συσχετίσεις (relations) εκ των οποίων η μία συνδέει τους μάγειρες με τα επεισόδια για την αποθήκευση των κριτών ανά επεισόδιο και η δεύτερη συσχετίζει τα επεισόδια με τους μάγειρες και τις εθνικές συνταγές. Ένας περιορισμός του E-R μοντέλου είναι ότι δεν μπορεί να εκφράσει συσχετίσεις μεταξύ συσχετίσεων.

Μία τέτοια δομή είναι αναγκαία στην περίπτωσή μας προκειμένου να διατηρούμε την πληροφορία της βαθμολογίας του κάθε κριτή σε κάθε μάγειρα ανά επεισόδιο. Ο καλύτερος τρόπος μοντελοποίησης αυτής της κατάστασης είναι η χρήση συνάθροισης. Όπως μπορούμε να παρατηρήσουμε, η συσχέτιση eval\_for ενώνει το αποτέλεσμα της συσχέτισης assignment και judges, προσθέτοντας το επιπλέον attribute evaluation.

1. **Σχεσιακό Σχήμα**

Ολοκληρώνοντας τον εννοιολογικό σχεδιασμό της βάσης προχωρήσαμε στον λογικό σχεδιασμό μετατρέποντας το E-R διάγραμμα σε σχεσιακό σχήμα (relational schema). Ακολουθήσαμε τις μεθόδους που διδαχθήκαμε στο μάθημα και τροποποιώντας την υλοποίησή μας όπου αυτό κρινόταν απαραίτητο, για την αποφυγή πλεονάζουσας πληροφορίας.

Όπως προαναφέρθηκε, κύριοι πίνακες της βάσης είναι αυτός των συνταγών και των μαγείρων. Κάθε άλλη οντότητα που σχετιζόταν με αυτές μετατράπηκε είτε σε επιπλέον attribute αυτών είτε σαν καινούριο πίνακα με foreign keys προς αυτούς. Επίσης οι μη δυαδικές μας σχέσεις μετατράπηκαν σε δυαδικές με χρήση βοηθητικών πινάκων. Όσον αφορά την υλοποίηση των περιορισμών των ελάχιστης και μέγιστης πληθυκότητας κάθε οντότητας που συμμετείχε σε σχέση, αυτό έγινε στο επίπεδο εισαγωγής των δεδομένων.

Το σχεσιακό μας μοντέλο συνοψίζεται στο παρακάτω σχήμα:

A screenshot of a computer program

Description automatically generated

1. **Ευρετήρια**

Τα ευρετήρια που δημιουργούνται αυτόματα λόγω των primary και foreign keys είναι:

1. Στον πίνακα Recipies, έχουμε το name, δηλαδή το όνομα της συνταγής στην οποία αναφερόμαστε (Primary Key).
2. Στον ίδιο πίνακα υπάρχει το attribute basic\_ingredient που είναι foreign key στον πίνακα Ingredients και έχουμε ένα attribute National\_Cuisine\_name που είναι foreign key στον πίνακα National\_Cuisine.
3. Ο πίνακας Ingredients έχει primary key το name του.
4. Ο πίνακας που περιέχει τα υλικά της κάθε συνταγής (Recipies\_has\_Ingredients) έχει foreign keys στον πίνακα των συνταγών και των υλικών.
5. Ο πίνακας Equipment έχει primary key το όνομα name του εξοπλισμού και ο Recipies\_has\_Equipment έχει foreign keys στα πίνακα του εξοπλισμού και των συνταγών.
6. Ο πίνακας των συμβουλών Tips έχει foreign key στον πίνακα των συνταγών.
7. Ο πίνακας Recipes\_has\_Steps στον οποίον αποθηκεύουμε τα βήματα κάθε συνταγής έχει primary key το ζεύγος Recipies\_name και step\_number και foreign keys στον πίνακα των συνταγών των βημάτων Steps.
8. Ο πίνακας των ετικετών etiquette έχει primary key το όνομα της ετικέτας Meal\_Type.
9. Ο πίνακας Recipies\_has\_Etiquette έχει primary key το ζεύγος Recipies\_name και Etiquette\_Meal\_Type και foreign key προς τον πίνακα των συνταγών και των ετικετών αντίστοιχα.
10. Ο πίνακας των μορφών γευμάτων Meal έχει primary key το Meal\_Form και ο πίνακας Meal\_has\_Recipies έχει primary key το ζεύγος Recipies\_name, Meal\_meal\_form και foreign keys προς τον πίνακα των συνταγών και των γευμάτων αντίστοιχα.
11. Ο πίνακας των θεματικών ενοτήτων Thematic\_Unit έχει primary key το όνομα name και ο πίνακας Recipies\_has\_Thematic\_Unit έχει primary key το ζεύγος Recipies\_name, Thematic\_Unit\_name και foreign keys στον πίνακα των συνταγών και των θεματικών ενοτήτων αντίστοιχα.
12. Ο πίνακας των εθνικών κουζίνων National\_Cuisine έχει primary key το όνομα name.
13. Ο πίνακας της εξειδίκευσης των μαγείρων Has\_specialisation έχει primary key την τριάδα National\_Cuisine\_name, Cook\_name, Cook\_surname και foreign keys προς τον πίνακα των Εθνικών κουζινών και των μαγείρων αντίστοιχα.
14. Ο πίνακας των μαγείρων Cook έχει primary key το ζεύγος name, surname και foreign key στον πίνακα της επαγγελματικής κατάρτισης Professional\_Expertise.
15. Ο πίνακας της επαγγελματικής κατάρτισης Professional\_Expertise έχει primary key το είδος type.
16. Ο πίνακας των επεισοδίων έχει primary key τον αριθμό των επεισοδίων episode\_number.
17. Ο πίνακας Assignment έχει primary key το σύνολο Αριθμός Επεισοδίου Episode\_episode\_number, το όνομα εθνικής κουζίνας National\_Cuisine\_name, το όνομα της συνταγής Recipies\_name, το ονοματεπώνυμο του μάγειρα, Cook\_name, Cook\_surname με foreign keys προς τους αντίστοιχους πίνακες.
18. Ο κύριος πίνακας της βάσης eval\_for έχει primary key τον συνδυασμό εθνικής κουζίνας, όνομα συνταγής, μάγειρα, κριτή και αριθμό επεισοδίου με foreign keys προς τα αντίστοιχα tables.

Πέρα από αυτά τα ευρετήρια που δημιουργούνται αυτόματα λόγων των primary και foreign κλειδιών, υλοποιήσαμε κάποια ευρετήρια για την βελτιστοποίηση της απόδοσης των ερωτημάτων επί της βάσης. Πιο συγκεκριμένα έχουμε:

1. Την ηλικία του μάγειρα age
2. Στον αριθμό των γραμμαρίων υδατανθράκων του πίνακα συνταγών
3. **DDL και DML**

Έχοντας έτοιμο το σχεσιακό μοντέλο, αξιοποιήσαμε το εργαλείο MySQLWorkbench για την ανάπτυξη του κώδικα της εφαρμογής σε MySQL. Πιο συγκεκριμένα, εκτός από τους ορισμούς των tables και των αντίστοιχων κλειδιών και ευρετηρίων τους, που θα περιγράψουμε πολύ αναλυτικά στην συνέχεια, δημιουργήσαμε τα διάφορα triggers που ελέγχουν τους περιορισμούς των δεδομένων της βάσης ή υπολογίζουν δυναμικά άλλα δεδομένα.

1. Ο υπολογισμός των θερμίδων της κάθε συνταγής υπολογίζεται από τις ποσότητες των υλικών που χρησιμοποιεί και τις θερμίδες αυτών ανά 100 γραμμάρια. Κάθε φορά που προσθέτουμε ένα στην βάση μας πληροφορία της μορφής (όνομα συνταγής, υλικό) στον πίνακα Recipes\_has\_ingredients, ένας trigger (after insert) αναλαμβάνει να υπολογίσει τον καινούριο αριθμών θερμίδων της συνταγής. Έτσι όταν θα έχουμε εισάγει όλα τα υλικά της συνταγής, θα έχουμε και τον τελικό αριθμό των θερμίδων της.
2. Κάθε φορά που εισάγουμε έναν νέο μάγειρα, η ηλικία του υπολογίζεται ως η διαφορά του timestamp την στιγμή της εισαγωγής και την ημερομηνίας γέννησής του (trigger before insert).
3. Στην βάση μας αρκεί να φορτώνουμε τους αριθμούς των επεισοδίων που θα θέλουμε να έχουμε και ένας trigger before insert στον πίνακα των επεισοδίων αναλαμβάνει να υπολογίσει τον αριθμό χρονιάς και σχετικά αριθμό επεισοδίου.
4. Κάθε φορά που προσθέτουμε κάποιον κριτή σε κάποιο επεισόδιο είμαστε σίγουροι ότι αυτό το επεισόδιο δεν θα έχει πάνω από 3 κριτές, χάρις την δημιουργία ενός trigger (before insert) που ελέγχει αν θα τηρείται ο αντίστοιχος περιορισμός. Αντίστοιχα υπάρχει trigger για να μην επιτρέπει την μη αποδεκτή διαγραφή κριτή.
5. Ο πίνακας όπου αποθηκεύουμε τις αναθέσεις μαγείρων, εθνικών κουζινών και συνταγών ανά επεισόδιο (Assignment) διαθέτει αντίστοιχα triggers με τον πίνακα των κριτών για αν διασφαλίζεται ότι τα επεισόδια κάθε χρονιάς θα είναι πράγματι 10.

Στο σημείο αυτό η βάση που δημιουργήσαμε είναι κατάλληλη για να πληρωθεί με δεδομένα (data loading). Παρόλα αυτά, ο φυσικός σχεδιασμός της αποφασίσαμε να περιλαμβάνει και stored procedures τα οποία θα δημιουργούν τις αναθέσεις του διαγωνισμού, παράγοντας ουσιαστικά τα «επεισόδιά»!

Mε την βοήθεια LLMs παράξαμε τα δεδομένα σχετικά με τις συνταγές και διασφαλίσαμε πως αυτά δεν είναι απλώς fake dump data αλλά αληθινές συνταγές που κανείς θα μπορούσε να μαγειρέψει. Το περιεχόμενων των αρχείων csv φορτώθηκαν στην βάση μέσω ενός Python Script που δημιουργήσαμε.

Καλώντας την ρουτίνα populate\_all δημιουργούμε 10 χρονιές του διαγωνισμού, με 10 επεισόδια η καθεμία, οι αναθέσεις των μαγείρων και των κριτών και η βαθμολόγηση των πιάτων. Η ρουτίνα αυτή καλεί τις παρακάτω ρουτίνες με την εξής σειρά:

1. Πρώτη καλείται η ρουτίνα for\_loop\_judges που καλεί 100 φορές (όσες και ο αριθμός των επεισοδίων) την ρουτίνα create\_judges. Με την σειρά της η ρουτίνα αυτή διαλέγει κάποιον τυχαίο μάγειρα, ελέγχει αν αυτός δεν έχει χρησιμοποιηθεί ως κριτής στα προηγούμενα δύο επεισόδια (κλήση ρουτίνας check\_judge\_not\_3\_in\_a\_row) και αν αυτός περάσει το τεστ, τον εισάγει στον πίνακα των κριτών. Έτσι εισάγονται 3 κριτές αν επεισόδια (300 κριτές συνολικά).
2. Δεύτερη καλείται η ρουτίνα create\_assignments που είναι και η πιο χρονοβόρα. Η ρουτίνα αυτή θα δημιουργήσει τις 10 αναθέσεις εθνικών κουζίνων σε μάγειρες και μετά σε συνταγές (συνολικά 1000 αναθέσεις). Κάθε φορά διαλέγεται τυχαία μία εθνική κουζίνα και από αυτή ένας μάγειρας στην οποία έχει ειδίκευση και μία συνταγή από αυτή την κουζίνα που ο μάγειρας ξέρει να μαγειρεύει. Καλούμε την ρουτίνα check\_valid\_assignment\_insert που ελέγχει αν ικανοποιούνται όλα τα foreign key constraints και είναι απλά μία επιπρόσθετη δικλείδα ασφαλείας ότι η ανάθεση αυτή είναι έγκυρη. Έπειτα ελέγχεται αν ο μάγειρας δεν έχει χρησιμοποιηθεί 3 φορές συνεχόμενα (κλήση ρουτίνας check cook not 3 in a row) και αντίστοιχα για την συνταγή (check\_recipe\_not\_3\_in\_a\_row) και για την εθνική κουζίνα (check\_national\_cuisine\_not\_3\_in\_a\_row).
3. Τελευταία καλείται η ρουτίνα populate\_eval\_for που αναλαμβάνει να κάνει το inner join μεταξύ των πίνακα των μαγείρων και των κριτών ανά επεισόδιο καθώς και να αναθέσει τυχαία τις βαθμολογίες τους σε κάθε προσπάθεια.

Για την δημιουργία της βάσης τοπικά αρκεί κανείς να τρέξει το αρχείο ddl της βάσης μας λινκ και έπειτα το αρχείο dml λινκ. Κατά αυτόν τον τρόπο η βάση μας είναι πλήρη και στην διάθεση των υφιστάμενων του διαγωνισμού για την ανάκτηση των πληροφοριών που τους ενδιαφέρει τρέχοντας τους κώδικες των αντίστοιχων SQL queries που τους παραδίδουμε.

1. **Authentication**

Πρόσβαση στην βάση δεδομένων πρέπει να έχουν τόσο οι διαχειριστές του διαγωνισμού όσο και οι διαγωνιζόμενοι. Δημιουργούμε έναν χρήση “admin” στον οποίο δίνουμε όλα τα δικαιώματα επί της βάσης, όπως τροποποίηση όλων των δεδομένων, δημιουργία αντιγράφου ασφαλείας και επαναφορά του συστήματος. Η είσοδος στην βάση γίνεται μετά την εισαγωγή του συνθηματικό “secure\_password”.

Για την υλοποίηση των λογαριασμών των μαγείρων, θα έπρεπε με την βοήθεια ενός trigger after insert, κάθε φορά που εισάγεται στην βάση ένας μάγειρας να δημιουργείται ένας αντίστοιχος λογαριασμός. Αυτός θα είχε πρόσβαση σε μονάχα 2 views, ένα που θα περιείχε τις προσωπικές του πληροφορίες και ένας δεύτερο με τις πληροφορίες των συνταγών που του έχουν ανατεθεί στον διαγωνισμό. Αυτή η δυναμική δημιουργία χρηστών δεν συνιστά καλή πρακτική και κάνει την βάση μας ευάλωτη σε SQL injections, έτσι το εργαλείο του MYSQLWorkbench εμπόδιζε αυτή την υλοποίηση. Εάν η εφαρμογή μας συνοδευόταν από UI τότε το authentication θα γινόταν σε αυτό το επίπεδο και το πρόβλημά μας θα πρακάμπτοταν. Επί της παρούσης αρκεστήκαμε στην δημιουργία ενός μονάχα μάγειρα χρήστη με τα αντίστοιχα views και δικαιώματα αλλαγών των δεδομένων που περιέχουν. Αξίζει να σημειωθεί πως έχουμε φροντίσει τα views να είναι updatable, δηλαδή να μην χρησιμοποιούν aggregate functions, group by ή having clauses, ενώσεις ή τομές, subqueries στο select και στο from και φυσικά joins.