

Data Mining Techniques – Εργασία 2

Ημερομηνία Παράδοσης: 1/4/2018

Η εργασία είναι ομαδική (ομάδες των 2 ατόμων).

Η 2^η εργασία αντιστοιχεί σε 20% του συνολικού βαθμού του μαθήματος.

Διδάσκοντας: Ι. Κωτίδης (kotidis@aueb.gr)

Βοηθός: Ι. Φιλιππίδου (filippidou@aueb.gr)

Περιγραφή

Σας δίνεται ένα αρχείο που περιλαμβάνει τα στοιχεία από 684 σκακιστικές παρτίδες που έχουν παιχθεί σε παγκόσμια τουρνουά σκακιού. Συγκεκριμένα το αρχείο περιλαμβάνει τα στοιχεία των παιχτών, τα τουρνουά, την χρονολογία, το αποτέλεσμα της παρτίδας αλλά και αναλυτικά όλες τις κινήσεις τις κάθε παρτίδας και τις θέσεις που προκύπτουν στην σκακιέρα με κάθε κίνηση. Το αρχείο με όλες τις παρτίδες υπάρχει στο e-class(chessData.txt) μαζί με την περιγραφή για κάθε στοιχείο του αρχείου (dataDescription.txt).

Σας ζητείτε να **μοντελοποιήσετε τα δεδομένα** ως ένα property graph σχεδιάζοντας τις κατάλληλες οντότητες και αναθέτοντας τα σχετικά labels, types και properties. Για την μοντελοποίηση θα χρειαστεί να μελετήσετε τα στοιχεία του αρχείου chessData.txt όπως αυτά περιγράφονται στο dataDescription.txt. Από τα δεδομένα αυτά θα πρέπει να αναπαρασταθούν όλα τα στοιχεία μιας παρτίδας σε κόμβους και ακμές αλλά και ειδικά όλες οι θέσεις και οι κινήσεις που αποτυπώνονται σε κάθε instance του αρχείου. Διευκρινίζεται ότι οι θέσεις κομματιών στην σκακιέρα είναι μοναδικές και μπορούν να προκύψουν με κινήσεις από πολλές παρτίδες. Στο σχεδιασμό θα πρέπει να συμπεριληφθούν σε κάθε τύπο κόμβου και ακμής μόνο τα στοιχεία που τον περιγράφουν και δεν θα πρέπει να υπάρχουν επαναλήψεις στοιχείων (ένα ίδιο property να εμφανίζεται και σε κόμβο και σε ακμή). Τέλος δεν θα πρέπει να συνδέονται κόμβοι με ακμές όταν αυτό δεν απαιτείται από το μοντέλο.

Στην συνέχεια θα πρέπει βάση του μοντέλου που σχεδιάσατε να **δημιουργήσετε μια βάση δεδομένων γράφων στο Neo4j και να φορτώσετε σε αυτή το σύνολο των δεδομένων** που σας δίνεται. Συγκεκριμένα θα πρέπει να δημιουργήσετε ένα μικρό πρόγραμμα σε μια γλώσσα της επιλογής σας, το οποίο θα δέχεται ως είσοδο το

αρχείο με τα δεδομένα που σας δίνεται και θα δημιουργεί νέα αρχεία (csv) κατάλληλα για είσοδο στο neo4j.

Τέλος μετά την εισαγωγή των δεδομένων στην βάση σας ζητείται **να γράψετε και να εκτελέσετε τα ακόλουθα ερωτήματα χρησιμοποιώντας την γλώσσα Cypher**. Για την επιτάχυνση της φόρτωσης αλλά και του χρόνου απόκρισης των ερωτημάτων θα μπορούσατε να δημιουργήσετε και τους κατάλληλους indexes στα properties του μοντέλου σας.

Ερωτήματα

1. Σε πόσες παρτίδες έχει προκύψει η θέση με
FEN: r1bqkbnrpprr1ppr2n51B2p34P35N2PPPP1PPPRNBQK2R
και τι ποσοστό αυτών έχει νικήσει ο λευκός.
2. Σε πόσες από τις παρτίδες που έχει προκύψει η θέση με
FEN: r1bqkbnrpprr1ppr2n51B2p34P35N2PPPP1PPPRNBQK2R νίκησαν τα λευκά, σε πόσες τα μαύρα και σε πόσες έγινε ισοπαλία.
3. Ποιο είναι το τουρνουά (event) που είχε τις περισσότερες παρτίδες και σε πόσες από αυτές έπαιζε ο “Karpon Anatoly” με λευκά ή μαύρα.
4. Ποιος παίκτης έχει παίξει τις περισσότερες φορές παρτίδες με άνοιγμα “Ruy Lopez”.
5. Πόσες και ποιοι είναι οι παίκτες των παρτίδων που έχουν την αλληλουχία κινήσεων “Nc6”, “Bb5”, “a6”.
6. Να εμφανίσετε τα στοιχεία της παρτίδας, το τουρνουά που παίχτηκε, τους παίκτες καθώς και όλες τις κινήσεις της παρτίδας με GameNumber:636 ταξινομημένες.
7. Να εμφανίσετε όλες τις παρτίδες που πέρασαν από την θέση με
Fen: r1bqkbnrpprr1ppr2n51B2p34P35N2PPPP1PPPRNBQK2R
και στην συνέχεια δεν παίχτηκε η κίνηση “a6”, επιπλέον να εμφανίσετε τις εναλλακτικές αυτές κινήσεις και το αποτέλεσμα της παρτίδας.

Παραδοτέο

Το παραδοτέο σας θα πρέπει να είναι ένα συμπιεσμένο αρχείο το οποίο θα αποστείλετε με mail στον βοηθό του μαθήματος και θα περιλαμβάνει:

1. Report.pdf:

- α. Ονοματεπώνυμο και ΑΜ των μελών της ομάδας σας.

b. Αναλυτική περιγραφή του μοντέλου δεδομένων που σχεδιάσατε με χρήση διαγράμματος και λεκτικής περιγραφής.

c. Περιγραφή της δομής των αρχείων που δημιουργήσατε για εισαγωγή στο neo4j, καθώς και τις εντολές που χρησιμοποιήσατε για την εισαγωγή των αρχείων στην βάση.

2. Το πρόγραμμα/script που φτιάξατε για την δημιουργία των αρχείων που χρησιμοποιεί το neo4j.

3. queries.cy: Αρχείο κειμένου με τα ερωτήματα όπως τα εκφράσατε στη γλώσσα Cypher.