

ΘΕΜΑ 1

“Παράλληλοι αλγόριθμοι για τριγωνισμό γραφημάτων”

Ο τριγωνισμός ενός γραφήματος έχει στόχο την εύρεση όλων των τριγώνων, δηλαδή την εύρεση όλων των τριάδων u,v,w έτσι ώστε να υπάρχουν οι ακμές (u,v) , (u,w) και (v,w) . Το πρόβλημα έχει πολλές εφαρμογές στην εξόρυξη δεδομένων από δεδομένα γραφημάτων. Στόχος της πτυχιακής είναι η μελέτη παράλληλων αλγορίθμων για την εύρεση των τριγώνων χρησιμοποιώντας multi-core αρχιτεκτονική. Εναλλακτικά, οι τεχνικές θα μπορούσαν να δοκιμαστούν και σε περιβάλλον Spark.

Απαιτήσεις: καλή γνώση C++ ή JAVA.

ΘΕΜΑ 2

“Αλγόριθμοι σε ροές γραφημάτων”

Μία ροή γραφήματος είναι ένα σύνολο ακμών που διαβάζονται με συγκεκριμένο τρόπο (συνήθως μία-προς-μία) και πρέπει χρησιμοποιώντας πολύ λίγο χώρο στη μνήμη να υπολογιστούν διάφορες ποσότητες για το γράφημα. Στην εργασία αυτή, θα μελετήσετε μερικές από τις τεχνικές αυτές και θα πραγματοποιήσετε και πειραματική μελέτη.

Απαιτήσεις: καλή γνώση C++ ή JAVA.

ΘΕΜΑ 3

“Τριγωνισμός γραφημάτων στο μοντέλο I/O”

Στην εργασία αυτή θα υλοποιηθούν αλγοριθμικές τεχνικές για την εύρεση όλων των τριγώνων ενός γραφήματος χρησιμοποιώντας το μοντέλο I/O. Στο μοντέλο αυτό, έχουμε περιορισμένη RAM και πρέπει να γίνουν πολλαπλά περάσματα στα δεδομένα ώστε να πάρουμε το τελικό αποτέλεσμα.

Απαιτήσεις: καλή γνώση C++ ή JAVA.

ΘΕΜΑ 4

“Μελέτη τεχνικών εύρεσης όμοιων αντικειμένων με χρήση SimHash”

Η τεχνική SimHash είναι μία μέθοδος παραγωγής fingerprints για αντικείμενα πολλών διαστάσεων. Σύμφωνα με την τεχνική αυτή, δύο αντικείμενα που διαφέρουν σε λίγα bits ως προς τα fingerprints ενδέχεται να διαφέρουν πολύ λίγο. Στην εργασία αυτή θα μελετήσουμε μεθόδους χρήσης της τεχνικής SimHash με σκοπό την αντιμετώπιση του προβλήματος του **plagiarism** σε μεγάλες συλλογές εγγράφων.

Απαιτήσεις: καλή γνώση C++ ή JAVA.

ΘΕΜΑ 5

“Υλοποίηση αλγορίθμων σε Apache Spark”

Το Spark αποτελεί σύστημα ανάπτυξης προγραμμάτων σε περιβάλλον cluster και μπορεί να χρησιμοποιηθεί από διάφορες γλώσσες όπως Python, Scala, Java ενώ μπορεί να συνεργαστεί πολύ καλά και με HDFS. Στην εργασία αυτή θα ασχοληθείτε με το Spark, και θα υλοποιήσετε μερικές τεχνικές για διαχείριση δεδομένων. Επίσης, θα υλοποιηθούν και τεχνικές στο σύστημα Spark SQL.

Απαιτήσεις: καλό θα είναι να υπάρχει επαφή με Java, Python ή Scala.

Παρακαλώ να δηλώσετε ενδιαφέρον μέχρι τις 31 Ιανουαρίου 2016, στέλνοντας και μία πρόχειρη αναλυτική βαθμολογία (papadopo@csd.auth.gr). Όλα τα θέματα απαιτούν καλές γνώσεις σε Αλγορίθμους, Δομές Δεδομένων και Βάσεις Δεδομένων. Τονίζεται ότι η πτυχιακή θα πρέπει να ολοκληρωθεί μέχρι το Σεπτέμβριο 2016 το αργότερο.

Απόστολος Ν. Παπαδόπουλος