

Προγραμματισμός & Συστήματα στον Παγκόσμιο Ιστό

Εργαστηριακή Άσκηση Ακαδημαϊκού Έτους 2021 - 2022

Θέμα: Πληθοποριστικά δεδομένα επισκεψιμότητας για έλεγχο διασποράς COVID-19

Έκδοση 1.0 (29/11/2021)

Στόχος

Στόχος της παρούσας εργασίας είναι η ανάπτυξη ενός συστήματος πληθοποριστικών δεδομένων, που αφορούν στην επισκεψιμότητα σημείων ενδιαφέροντος και τη διαχείριση πιθανής επαφής με κρούσμα COVID-19.

Λειτουργικές Προδιαγραφές

Η διαχείριση της πανδημίας από την COVID-19 και των συνεπειών της αποτελεί πρόκληση σε πολλές χώρες του κόσμου. Για τον έλεγχο της πιθανής διασποράς μπορούν να χρησιμοποιηθούν πληθοποριστικά δεδομένα επισκεψιμότητας που ήδη συλλέγονται από παρόχους, όπως η Google και αφορούν στην πυκνότητα επισκέψεων σε γεωγραφικά σημεία ενδιαφέροντος (POIs), όπως καταστήματα, εστιατόρια, υπηρεσίες κατά τη διάρκεια της ημέρας και της εβδομάδας (popular times).

Η πυκνότητα επισκέψεων μπορεί να λειτουργήσει ως δείκτης της δημοφιλίας ενός μέρους, αλλά και ως παράγοντας κινδύνου για πιθανή διασπορά COVID-19. Παράλληλα, αποτελεί μόνο εκτίμηση με βάση τα δεδομένα γεωγραφικού εντοπισμού και δεν εξαρτάται από τον πραγματικό αριθμό επισκεπτών σε ένα χώρο. Θα ήταν επομένως χρήσιμο να δίνεται η δυνατότητα σε έναν επισκέπτη να συνεισφέρει την προσωπική του εκτίμηση για τον πραγματικό αριθμό των επισκεπτών.

Διατηρώντας με ανώνυμο τρόπο την πληροφορία των επισκέψεων, μπορεί να γίνει συνδυασμός με καταχώρηση πιθανού κρούσματος, ώστε να υπάρχει ενημέρωση σε άλλους χρήστες αν βρέθηκαν στον ίδιο χώρο με δηλωμένο κρούσμα. Έτσι, σκοπός της εργασίας είναι να κατασκευαστεί ένα σύστημα πληθοποριστικής συλλογής δεδομένων επισκεψιμότητας και δήλωσης πιθανού κρούσματος με σκοπό την παροχή πληροφοριών επισκεψιμότητας σε σημεία ενδιαφέροντος, αλλά και την ενημέρωση για πιθανή επαφή με δηλωμένο κρούσμα, ως μέσο για έλεγχο της διασποράς. Στο σύστημα υπάρχουν δύο τύποι χρηστών: Διαχειριστής και Χρήστης.

Χρήστης

Η αλληλεπίδραση με το χρήστη γίνεται μέσω προσαρμοστικού ιστοτόπου που επιτρέπει την πρόσβαση μέσω σταθερού υπολογιστή ή κινητού τηλεφώνου, και έχει τις εξής δυνατότητες:

- 1) **Εγγραφή στο σύστημα.** Ο χρήστης εγγράφεται και αποκτά πρόσβαση στο σύστημα επιλέγοντας κάποιο **username & password** της αρεσκείας του, και παρέχοντας το **email** του. Το password πρέπει να είναι **τουλάχιστον 8 χαρακτήρες** και να περιέχει τουλάχιστον **ένα κεφαλαίο γράμμα, ένα αριθμό** και **κάποιο σύμβολο** (π.χ. # \$ * & @).
- 2) **Αρχική απεικόνιση χάρτη.** Στον χρήστη εμφανίζεται ένας χάρτης με επίκεντρο τη γεωγραφική θέση που βρίσκεται τη στιγμή της πρόσβασης. Η θέση του χρήστη εμφανίζεται με κατάλληλη οπτική ένδειξη στο χάρτη. Ο χάρτης πρέπει να έχει κατάλληλο επίπεδο **οπτικής μεγέθυνσης** ώστε να

καλύπτει μια περιοχή ακτίνας **όχι μικρότερης από 5χλμ** από τη θέση του χρήστη. Στον χάρτη δεν εμφανίζονται κατά την αρχική πρόσβαση markers που αντιστοιχούν σε POIs της βάσης δεδομένων.

- 3) **Αναζήτηση σημείων ενδιαφέροντος.** Μέσω πλαισίου ελεύθερης αναζήτησης μπορεί να αναζητήσει σημείο ενδιαφέροντος (POI) από τη **βάση δεδομένων του συστήματος**, και το οποίο να είναι εντός του **ορατού εύρους του** χάρτη για πιθανή επίσκεψη (εστιατόριο, καφέ, υπηρεσία κλπ). Τα αποτελέσματα της αναζήτησης αυτής εμφανίζονται **ως markers**. Κάνοντας κλικ σε έναν marker εμφανίζεται σε **αναδυόμενο pop-up** εκτίμηση της επισκεψιμότητας στο μέρος αυτό **με βάση τα δεδομένα που συλλέχθηκαν, για τις επόμενες 2 ώρες από την τρέχουσα**. Επίσης εμφανίζεται **ο μέσος αριθμός των επισκεπτών** με βάση τις καταχωρήσεις άλλων χρηστών, **αν υπάρχουν ως και 2 ώρες πριν την τρέχουσα χρονική στιγμή (βλ. επόμενο σημείο)**. Οι markers θα πρέπει να έχουν κατάλληλη χρωματική ένδειξη ανάλογα με την **εκτιμώμενη επισκεψιμότητα** (**0-32% Πράσινο, 33-65% Πορτοκαλί, 66%+ Κόκκινο**).
- 4) **Καταχώρηση επίσκεψης.** Για τα POIs που βρίσκονται εντός **20 μέτρων από την τοποθεσία του χρήστη**, εμφανίζεται, στο ίδιο pop-up με το (2), **η επιλογή καταχώρησης της επίσκεψης του χρήστη**. Εφόσον ο χρήστης συναινέσει, αποθηκεύεται στη ΒΔ του συστήματος (server) η επίσκεψή του στο συγκεκριμένο σημείο με τρέχον **timestamp**. Παράλληλα, έχει τη δυνατότητα να **καταχωρήσει μια εκτίμησή του για τον αριθμό των ατόμων στο σημείο (ακέραιος αριθμός)**.
- 5) **Δήλωση κρούσματος.** Ο χρήστης έχει τη δυνατότητα να δηλώσει στην εφαρμογή αν ο ίδιος αποτελεί **κρούσμα COVID-19** και τότε διαγνώστηκε (**ημερομηνία**). **Εκ νέου δήλωση του ίδιου χρήστη δεν έχει αποτέλεσμα, εκτός αν έχουν παρέλθει 14 ημέρες** από την αρχική ημερομηνία διάγνωσης.
- 6) **Πιθανή επαφή με κρούσμα.** Ο χρήστης μπορεί να δει αν έχει έρθει σε επαφή με κρούσμα τις τελευταίες 7 μέρες. Εμφανίζεται λίστα με τα σημεία ενδιαφέροντος που ο χρήστης επισκέφτηκε, συμπεριλαμβανομένης ημερομηνίας και ώρας, και για τα οποία έχει καταγραφεί κρούσμα που: α) βρισκόταν στο ίδιο σημείο με το χρήστη σε ένα εύρος ± 2 ωρών και β) έχει διαγνωσθεί ως κρούσμα εντός 7 ημερών από την επίσκεψη.
- 7) **Επεξεργασία προφίλ.** Ο χρήστης μπορεί να αλλάξει το username/password και να δει το αναλυτικό ιστορικό των επισκέψεών του και δηλώσεων κρούσματος που έχει κάνει.

Διαχειριστής

Ο Διαχειριστής αποκτά πρόσβαση στο σύστημα με **σταθερό υπολογιστή**, μέσω κατάλληλου μηχανισμού **username / password**. Ο διαχειριστής δεν εγγράφεται στο σύστημα αλλά δημιουργείται ως οντότητα απευθείας στην σχετική βάση δεδομένων. Κατά την είσοδό του στο σύστημα έχει τις εξής δυνατότητες:

- 1) **Ανέβασμα, ενημέρωση και διαγραφή δεδομένων.** Ο διαχειριστής έχει τη δυνατότητα να **φορτώσει** δεδομένα σημείων ενδιαφέροντος (POIs) και της επισκεψιμότητάς τους που πληρούν τις προϋποθέσεις συμμόρφωσης με το πρότυπο δίνεται (JSON). Τα οποία αποθηκεύονται στη ΒΔ. Μπορεί επίσης να επικαιροποιήσει τα δεδομένα αυτά, χωρίς όμως να δημιουργούνται διπλοεγγραφές. Μπορείτε να θεωρήσετε ως ημερομηνία άντλησης των δεδομένων την ημερομηνία τελευταίας τροποποίησης του αρχείου στο δίσκο. Επίσης μπορεί να επιλέξει τη διαγραφή όλων των δεδομένων.
- 2) **Απεικόνιση Στατιστικών.** Ο διαχειριστής βλέπει σε μία σελίδα κατάλληλη πληροφορία, σε πίνακες ~~ή~~ και γραφήματα τα οποία απεικονίζουν:
 - a. Συνολικός αριθμός επισκέψεων που έχουν καταγραφεί.
 - b. Συνολικός αριθμός κρουσμάτων που έχουν δηλωθεί.
 - c. Συνολικός αριθμός επισκέψεων από ενεργά κρούσματα: Η επίσκεψη θα πρέπει να έχει γίνει έως και 7 ημέρες πριν την ημερομηνία διάγνωσης κρούσματος και μέχρι 14 ημέρες μετά.
 - d. Κατάταξη των κατηγοριών σημείων ενδιαφέροντος που υπάρχουν στη ΒΔ του συστήματος, με βάση το συνολικό πλήθος των επισκέψεων, όπως καταχωρείται από τους χρήστες.

- e. Κατάταξη των κατηγοριών σημείων ενδιαφέροντος που υπάρχουν στη ΒΔ του συστήματος, με βάση τον αριθμό κρουσμάτων που είχαν προσέλθει σε αυτά, έως και 7 ημέρες πριν την ημερομηνία διάγνωσης κρούσματος και μέχρι 14 ημέρες μετά.
- f. Εμφάνιση διαγράμματος ανά ημέρα, όπου:
 - i. Επιλέγεται ένα προκαθορισμένο εύρος (συγκεκριμένη εβδομάδα ή συγκεκριμένος μήνας).
 - ii. Επιλέγεται ~~με~~ φίλτρο η δυνατότητα εμφάνισης ενός ή περισσότερων από τα παρακάτω:
 - 1. Το πλήθος των επισκέψεων που δηλώνουν οι χρήστες για κάθε ημέρα.
 - 2. Το πλήθος των επισκέψεων από διαγνωσμένα κρούσματα για κάθε ημέρα
- g. Εμφάνιση διαγράμματος ανά ώρα, όπου:
 - i. Επιλέγεται μια συγκεκριμένη ημερομηνία.
 - ii. Επιλέγεται με φίλτρο η δυνατότητα εμφάνισης ενός ή περισσότερων από τα παρακάτω:
 - 1. Η μεταβολή του αριθμού επισκέψεων ανά ώρα.
 - 2. Η μεταβολή του αριθμού επισκέψεων από δηλωμένο κρούσμα ανά ώρα.

Περιορισμοί

1. Ομάδες 3 (τριών) το πολύ ατόμων.
2. Ο πυρήνας του παραδοτέου σας θα πρέπει υποχρεωτικά να υλοποιείται με τεχνολογίες που έχουν διδαχθεί στο μάθημα (PHP, JavaScript, Node.JS, HTML).
3. Η εμφάνιση και η λειτουργικότητα της εφαρμογής αξιολογείται.

Παραδοτέα

1. Συνοπτική αναφορά που θα περιλαμβάνει:
 - a. Τον σχεδιασμό της βάσης (Entity-Relationship Diagram για σχεσιακές βάσεις, περιγραφή του βασικού μοντέλου δεδομένων για μη-σχεσιακές βάσεις).
 - b. Τις ρυθμίσεις που κάνατε στον server σας για να αξιοποιήσετε κατάλληλα τις δυνατότητες των κρυφών μνημών (caches, π.χ. θέτοντας TTLs για τα διαφορετικά είδη αρχείων). Αναφέρετε τις σχετικές παραμετροποιήσεις που κάνατε στο configuration του server, ή μέσω .htaccess, ή in-line ορισμό κεφαλίδων απόκρισης και δώστε μερικά screenshots από ~~τα developer tools~~ του browser που χρησιμοποιείτε, ώστε να δείξετε ότι οι ρυθμίσεις που κάνατε είχαν αποτέλεσμα. Στηρίξτε τις σχετικές επιλογές που αφορούν τους κανόνες που υλοποιήσατε ή τις τιμές που επιλέξατε για διάφορες παράμετρους, με σχετικές παραπομπές σε τεχνικά άρθρα ή επιστημονική βιβλιογραφία.
 - c. Αν κατεβάσατε δεδομένα και τα μοιραστήκατε στο κοινό αποθετήριο (μαζί με τον αποδεικτικό κωδικό ανεβάσματος)
 - d. Τη λογική με την οποία υλοποιήθηκε το script δημιουργίας επισκέψεων χρηστών.
2. Τον πηγαίο κώδικα και ένα export της ΒΔ
3. Όλα τα ανωτέρω (1) και (2) υποβάλλονται σε ένα συμπιεσμένο αρχείο στο eclass.

Χρήση τεχνολογιών

- Θα χρησιμοποιήσετε αποκλειστικά open-source τεχνολογίες για τη ΒΔ (MySQL, PostgreSQL, MongoDB), τους χάρτες (προτείνεται η βιβλιοθήκη Leaflet), τα γραφήματα (προτείνεται η βιβλιοθήκη chart.js).
- Η φόρτωση δεδομένων από τη ΒΔ θα πρέπει να γίνει χρησιμοποιώντας αποκλειστικά τεχνικές AJAX (προσοχή: όχι PHP για τη δημιουργία Javascript κώδικα).
- Προσοχή στη βάση δεδομένων: Χρησιμοποιήστε κατάλληλα indexes στους πίνακες ώστε να επιταχύνονται τα queries. Επίσης φροντίστε ώστε η είσοδος δεδομένων στη βάση να γίνεται με bulk inserts (π.χ. 1000 εγγραφές μαζί) και όχι μια-μια εγγραφή, για να επιταχύνεται η διαδικασία.

Παράρτημα Τεχνικών Οδηγιών

Δεδομένα σημείων ενδιαφέροντος

Ως σημείο αφετηρίας για την εργασία σας, θα βρείτε στον αποθετήριο δεδομένων της άσκησης (δείτε πιο κάτω) ένα μικρό σύνολο σημείων ενδιαφέροντος με τα δεδομένα δημοφιλίας τους. Τα δεδομένα έχουν αντληθεί με χρήση ενός crawler σε Python που μπορείτε να βρείτε εδώ <https://github.com/m-wrzs/populartimes>

Στο ανωτέρω σύνδεσμο δίνεται και ένα πλήρες παράδειγμα της δομής (format) των δεδομένων σε JSON.

Για να μπορέσετε να δοκιμάσετε πιο αποτελεσματικά την εργασία σας με πληθώρα δεδομένων, μπορείτε προαιρετικά να κατεβάσετε πρόσθετα δεδομένα εκτελώντας τον crawler σύμφωνα με τις οδηγίες που υπάρχουν στο github.

Αν κατεβάσετε δικά σας σύνολα δεδομένων, τότε σας ενθαρρύνουμε να τα μοιραστείτε με τους υπόλοιπους συμμετοχτές σας, ανεβάζοντάς τα στο κοινό αποθετήριο που έχει δημιουργηθεί για το σκοπό αυτό στον ακόλουθο σύνδεσμο:

<http://150.140.142.67:8083/web21/>

Για τους σκοπούς της άσκησης παρακαλούμε να χρησιμοποιηθεί αποκλειστικά ο συγκεκριμένος crawler ώστε τα δεδομένα στο αποθετήριο να έχουν κοινή δομή και να είναι χρησιμοποιήσιμα από όλους.

Για να ανεβάσετε δεδομένα στο κοινό αποθετήριο θα πρέπει:

- Τα δεδομένα που θα κάνετε upload θα πρέπει να είναι σε μορφή zipped JSON (.zip)
- Κάθε zip που ανεβάζετε μπορεί να περιέχει περισσότερα από 1 JSON αρχεία (π.χ. αν εκτελέσετε τον crawler σε διαφορετικές χρονικές στιγμές, ή για διαφορετικές τοποθεσίες).
- Μαζί με τα JSON αρχεία θα πρέπει να συμπεριλάβετε ένα txt αρχείο (readme.txt) το οποίο περιέχει πληροφορίες για κάθε αρχείο στην παρακάτω μορφή (μια γραμμή για κάθε αρχείο)
 - [όνομα αρχείου] \t [ημερομηνία συλλογής]
 - Παράδειγμα:
 - file1.json 2021-11-23 13:03:49
 - file1.json 2021-11-24 15:08:43
- Με το upload θα σας εμφανιστεί ένας μοναδικός κωδικός (πιστοποιητικό), τον οποίο θα πρέπει να καταγράψετε και να τον αναφέρετε στην αναφορά του project σας. Μετά το upload, το αρχείο σας θα είναι διαθέσιμο προς χρήση και από τις υπόλοιπες ομάδες.

Δεδομένα χρήσης συστήματος και επισκέψεων χρηστών

Για να γίνει αποτελεσματικά η δοκιμή της εργασίας, θα πρέπει να προσομοιώσετε την ύπαρξη κάποιων χρηστών στο σύστημα. Θα πρέπει να δημιουργήσετε τουλάχιστον 5 χρήστες με τουλάχιστον 100 επισκέψεις συνολικά (π.χ. 5x20) οι οποίες θα πρέπει να επικαλύπτονται χρονικά κατά ένα μέρος (π.χ. να έχουν συμβεί όλες εντός του ίδιου μήνα). Ένα ποσοστό (π.χ. 20%) αυτών των επισκέψεων θα πρέπει να έχουν δημιουργηθεί από διαγνωσμένο κρούσμα.

Η ανωτέρω διαδικασία μπορεί να γίνει χειροκίνητα. Ωστόσο συστήνεται να φτιάξετε κάποιο script σε PHP ή JavaScript, το οποίο να δημιουργεί αυτόματα λογαριασμούς χρηστών και τις επισκέψεις τους, ώστε να δημιουργούνται εγγραφές στη βάση χωρίς ανθρώπινη παρέμβαση. Το script αυτό μπορεί να καλείται από τη διεπαφή του διαχειριστή και να δέχεται σαν είσοδο:

- Το πλήθος των χρηστών N, ώστε να δημιουργούνται οι λογαριασμοί τους με user name "UserX" όπου X=[1..N], και password "userXpass".
- Μια χρονική περίοδο (ημερομηνία έναρξης, τέλους) που θα αφορά τη δημιουργία των επισκέψεων αυτών των χρηστών (π.χ. 01-11-2021 με 31-12-2021)

Και στη συνέχεια, το script δημιουργεί καταχωρήσεις επίσκεψης με **τυχαίο αλλά σχετικά ρεαλιστικό τρόπο**, π.χ. μια λογική θα μπορούσε να είναι:

- Κάθε ημέρα ένα τυχαίο αριθμό επισκέψεων (π.χ. μεταξύ 5-10)
- Η πρώτη επίσκεψη της ημέρας γίνεται σε τυχαίο σημείο από όλα τα διαθέσιμα.
- Κάθε επόμενη επίσκεψη για την ίδια ημέρα να είναι σε κάποιο τυχαίο σημείο (POI) από το σύνολο των σημείων (POIs) που βρίσκονται σε απόσταση το πολύ K χιλιομέτρων (π.χ. $K=2$, με τυχαία επιλογή του $K=[1...4]$) από την προηγούμενη.
- Κάθε επίσκεψη να απέχει χρονικά από την προηγούμενη ανάλογα με την απόσταση που τις χωρίζει (θεωρώντας πως ο χρήστης κινείται με τα πόδια, με μέση ταχύτητα 3χλμ την ώρα), συν ένα πρόσθετο τυχαίο χρόνο παραμονής (από 10 ως 120 λεπτά).

Σημειώνεται ότι η εθελοντική προσφορά δεδομένων ΔΕΝ προσμετράται στη βαθμολογία. Η χρήση ενός script για τη δημιουργία δεδομένων επίσκεψης αντιστοιχεί σε bonus +1 μονάδας στον τελικό βαθμό (με μέγιστο το 10).