

Rozwiązanie zadania Betacom

Parametry wejściowe zapytania:

carId – id pojazdu dla którego szukamy gpsId,

recordsToExamineCount - Ilość rekordów z tablicy routes do analizy,

accuracy – dokładność (0-1) Ile rekordów musi pasować aby uznać że dany gpsId przynależy do carId

stopTime – szacunkowy czas postoju pojazdu

gpsOffset – odchyłka GPS (ile dwa koordynaty mogą różnić się od siebie aby uznać je za jedną lokalizację)

1. Wybranie dla którego **carId** szukamy **gpsId**,
2. Określenie od jakiej daty (**startGpsTracesTimestamp**) zaczynają się timestampy w tablicy gps (rekordy są tylko z danego dnia),
3. Znalezienie liczby **recordsToExamineCount** rekordów w tablicy routes, które mają timestampy większe od **startGpsTracesTimestamp** i mają interesujące nas (**carId**). (lista **routesAfterStartTime**)

Komentarz:

Zapisujemy listę kilku kolejnych rekordów aby nie szukać ich potem po tablicy routes (kosztem pamięci). Ilość rekordów **RecordsCount** może być konfigurowalna ponieważ gdy damy małą ilość rekordów istnieje szansa że nie trafimy pasującego rekordu z tablicy gps. Istnieje również ryzyko że dwa pojazdy w ciągu dnia będą w jednym sklepie obok siebie na wyładunku. Z tych powodów warto wytypować więcej rekordów do późniejszej analizy.

4. Dla każdego rekordu (**routeAfterStart**) z listy **routesAfterStartTime** znalezienie w tablicy „gps” rekordów spełniających założenia :
 - Timestamp nowszy od **routeAfterStart.timestamp**,
 - Timestamp mniejszy niż wartość konfigurowalna **routeAfterStart.timestamp + stopTime**

Komentarz:

Pojazd zatrzymuje się na wyładunek więc istnieje duża szansa że przez chwilę będzie w jednym miejscu.

5. Pogrupowanie wyników po gpsId.(lista **groupByIdTracesInTimeRangeAfterArrivalTime**)
6. Dla każdej grupy:
Porównanie każdego koordynatu z odpowiadającym mu koordynatem obiektu **routeAfterStart**.
Koordynaty porównywane z uwzględnieniem odchyłki **gpsOffset**.
7. Stworzenie dwóch kolekcji słownikowych :
 - **D1** (gpsId,matchingRouteGpsTrace) – liczba pasujących rekordów dla każdego gpsId
 - **D2** (gpsId,allGpsTracesFromRange) – liczba wszystkich rekordów z gps znajdujących się w danym przedziale czasu (punkt 4) dla każdego gpsId
8. Określenie dla każdego elementu z tablicy D1 stosunku liczby pasujących rekordów do wszystkich rekordów. Zapisanie każdego wyniku w kolekcji słownikowej :
 - **D3** (gpsId, resultOfMatching) – rezultaty dopasowań dla każdego gpsId,
9. Wybranie z kolekcji **D3** wyniku o najlepszym rezultacie. W przypadku gdy dwa wyniki są jednakowe odrzucenie całego wyszukiwania jako niejednoznacznego.
10. Sprawdzenie czy wynik jest większy od założonej dokładności,
11. Zwrócenie wyniku gpsId