

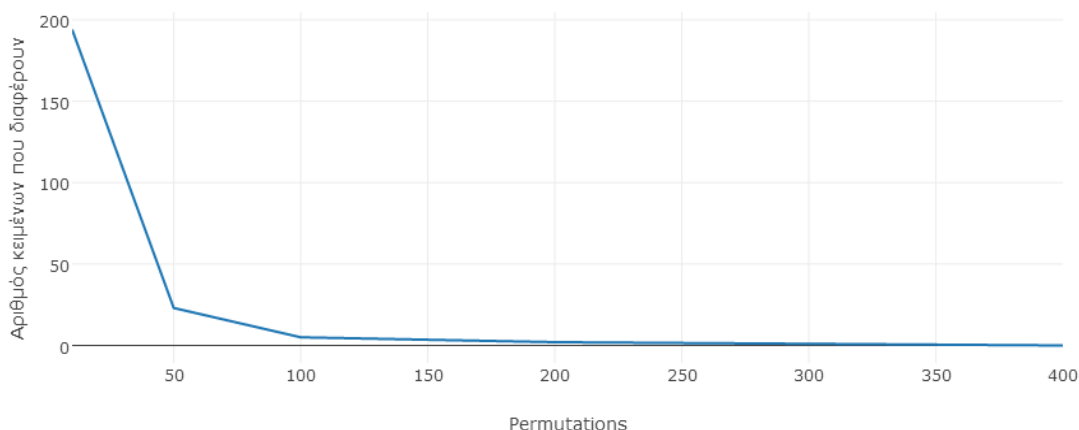
## ΑΠΑΝΤΗΣΕΙΣ ΕΡΩΤΗΜΑΤΩΝ

Ερώτημα 4: Μετά από μέτρηση με τη χρήση της `System.currentTimeMillis()` πρόκυψαν τα παρακάτω αποτελέσματα μετά από 5 εκτελέσεις:

Permutations	Αριθμός κειμένων που διαφέρουν	Χρόνος ακριβούς Jaccard	Χρόνος Minhash Jaccard
10	194	597762 ms	1 ms
50	23	603056 ms	2 ms
100	5	603563 ms	4 ms
200	2	609447 ms	5 ms
400	0	629871 ms	8 ms

Ερώτημα 5: Παρατηρείται ότι όσο αυξάνεται ο αριθμός των *permutations* τόσο μειώνεται ο αριθμός των κειμένων που διαφέρουν. Επίσης αυξάνεται ο χρόνος τόσο για την *exact Jaccard* όσο και για την *approximate Jaccard*. Επιπλέον, ο χρόνος *ακριβούς Jaccard* είναι κατά πολύ πιο αργός σε σχέση με τον χρόνο *Minhash Jaccard*.

Ερώτημα 6: Η πιο κατάλληλη τιμή των *permutations* είναι το 100. Αν επιλεγόταν το 10 ο αριθμός των κειμένων που διαφέρουν θα ήταν πάρα πολύ μεγάλος, ενώ αν επιλεγόταν το 400 ή το 200 θα μας έδινε πολύ μικρό αριθμό διαφορετικών κειμένων. Επίσης παρατηρείται ότι από 10 *permutations* σε 50 έχουμε πολύ μεγάλη μείωση από 194 → 23 ενώ μετά το 100 είναι πιο ομαλή η μείωση του αριθμού των διαφορετικών κειμένων από 5 → 2. Οπότε το 100 επιλέγεται σαν πιο κατάλληλη τιμή για *permutations* γιατί από εκεί και πέρα η μείωση είναι πιο ομαλή.



Ερώτημα 8: Μετά από μέτρηση με τη χρήση της *System.currentTimeMillis()* πρόκυψαν τα παρακάτω αποτελέσματα μετά από 5 εκτελέσεις για *permutations 100*:

<b>Bands</b>	<b>Χρόνος LSH + MinHash Jaccard</b>	<b>Αριθμός False Positives</b>
25	222271 ms	200
10	103461 ms	80
2	48696 ms	28
1	2364 ms	4

Ερώτημα 9: Παρατηρείται ότι όσο μειώνεται ο αριθμός των *bands* τόσο μειώνεται και ο χρόνος *LSH + MinHash Jaccard*(μείωση σχεδόν στο μισό του χρόνου κάθε φορά που μειώνεται το *bands*), καθώς επίσης μειώνεται και ο αριθμός των *false positives*. Σε σχέση με τον προηγούμενο χρόνο *MinHash Jaccard*, ο χρόνος *LSH + MinHash Jaccard* είναι κατά πολύ πιο αργός.