IS624 - Assignment 1

James Quacinella

06/13/2015

Exercise 1.2.1

Using the information from Section 1.2.3, what would be the number of suspected pairs if the following changes were made to the data (and all other numbers remained as they were in that section)?

(a) The number of days of observation was raised to 2000.

(b) The number of people observed was raised to 2 billion (and there were therefore 200,000 hotels).

(c) We only reported a pair as suspect if they were at the same hotel at the same time on three different days.

Answer

Original formaula: $5 * 10^{17} * 5 * 10^5 * 10^{-18} = 250000$

(a) If the number of days is changed to 2000, then the approximate $5 * 10^5$ would be higher, since it'll now be 2000 choose 2, which is approximately 2*10ˆ6. The new result would be:

$5 * 10^{17} * 2 * 10^6 * 10^{-18} = 1000000$

The number of pairs would increase by a factor of 4.

(b) Raising the number to 2 billion affects the chance of visiting the same hotel. Before, that was 10^{-9}. However, with 200,000 hotels, we have $\frac{.0001}{2*10^9}$ which is $5 * 10^{-10}$. The chance that they will visit the same hotel on two different given days is now $5 * 10^{-20}$.

This also affects the the number of pairs of people, which is now approximately nˆ2/2 where n = $2 * 10^9$, which evaluates to 4*10ˆ18 / 2 = 2^{10}18.

Re-evaluating everything, we now have:

$2 * 10^{18} * 5 * 10^5 * 5 * 10^{-20} = 50000$

(c) Three consecuative days means that the calculation 1000 choose 2 should now be 1000 choose 3, which evaluates to approximately $1.7 * 10^9$. The result should now be:

$5 * 10^{17} * 1.7 * 10^9 * 10^{-18} = 8.5 * 10^8 = 850000000$

Exercise 1.3.2

Suppose there is a repository of ten million documents, and word w appears in 320 of them. In a particular document d, the maximum number of occurrences of a word is 15. Approximately what is the TF.IDF score for w if that word appears (a) once (b) five times?

Answers

We are given $N = 10^7$ and $n_i = 320$. Also, we know $max_k f_{kj} = 15$. Therefore we know that the IDF score is $IDF_i = log_2(N/n_i) = log2(10^7/320) = 14.931$. Only the TF score changes.

(a) $TF_{ij} = \frac{f_{ij}}{max_k f_{kj}} = 1/15$

Therefore $TFIDF = TF_{ij} * IDF_i = \frac{14.931}{15} = 0.9954$.

(b) $TF_{ij} = \frac{f_{ij}}{max_k f_{kj}} = \frac{5}{15} = \frac{1}{3}$

Therefore $TFIDF = TF_{ij} * IDF_i = \frac{14.931}{3} = 4.977$.

Question

Use the Taylor expansion of e x to compute, to three decimal places: (a) $e^{1/10}$ (b) $e^{-1/10}$ (c) e^2 .

Answers

(a) The taylor series expansion for $e^{1/10}$ looks like:

$$e^{1/10} = 1 + (1/10) + \frac{(1/10)^2}{2} + \frac{(1/10)^3}{6} + ...$$

which evaluates to 1.105167, or 1.105 to three decimal places.

(b) Instead of evaluating, we can take the reciprocal of the above answer, which ends up being .90497, or .905 to three deciaml places. The taylor expansion would have alternating signs since the value is negative.

(c) The taylor series expansion for e^2 looks like:

$$e^2 = 1 + (2) + \frac{(2)^2}{2} + \frac{(2)^3}{6} + \frac{(2)^4}{24} + \frac{(2)^5}{120} + \frac{(2)^6}{720} + + \frac{(2)^7}{7!} + \frac{(2)^8}{8!}$$
$$= 1 + 2 + 2 + 1.33333 + 0.6666667 + 0.2666667 + 0.08888889 + 0.02539683 + 0.006349206$$

which is approximately 7.387 to three decimals.