Expected Overlap of Words in Context (EOWC)

Probability that a randomly picked word from d1 is wi

Count of word wi in d1

$d1=(x_1, \ldots x_N)$ $x_i = c(w_i, d1)/|d1|$

$d2=(y_1, \ldots y_N)$ $y_i = c(w_i, d2)/|d2|$

Total counts of words in d1

$$Sim(d1,d2)=d1.d2= x_1 y_1 + \ldots + x_N y_N = \sum_{i=1}^{N} x_i y_i$$

Probability that two randomly picked words from d1 and d2, respectively, are identical.