Application of Hashing in NLP

Write Python program for the problem below.

You are given a set of text documents and your goal is to find the most similar text document with respect to a query text. The similarity is computed between two text documents, d_1 and d_2 as follows:

$$similarity(d_1, d_2) = \sum_{w \in V} p(w|d_1) \log(\frac{p(w|d_1)}{p(w|d_2)})$$

Where V is the vocabulary (all the words), $p(w|d)$ is the probability of the word w in d.

$$p(w|d) = \frac{(number\ of\ times\ w\ present\ in\ d) + 1}{(total\ number\ of\ words\ in\ d) + 2}$$

Your objective is to represent each document in a hash table and when the query comes in, search for each query word in each of the documents to find the probability. Finally compute the similarity score.