1. Cosine Similarity:

$$\cos(\theta) = \frac{\vec{d_i} \cdot \vec{d_j}}{\|\vec{d_i}\|\|\vec{d_j}\|}$$

where $\vec{d_i}$, $\vec{d_j}$ are bag-of-words representations (each entry stores the term frequency of a particular word) of document i and j, respectively.

[1] Singhal, Amit (2001). "Modern Information Retrieval: A Brief Overview". Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24 (4): 35–43.

2. Jaccard Similarity

$$J(S_i, S_j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|}$$

where S_i and S_j are the **sets** of words (a binary indicator) in document i and j, respectively. e.g. suppose document 1 is "aacbc", and there are in total four labels: 'a', 'b', 'c', 'd', then $\vec{d_1} = (2, 1, 2, 0)$ but S_i can be represented as $(1, 1, 1, 0)$.

[2] Tan, Pang-Ning; Steinbach, Michael; Kumar, Vipin (2005), Introduction to Data Mining, ISBN 0-321-32136-7

3. Minimum Edit Distance:

$$\textbf{MED} = D + I + 2S$$

where D is the number of deletions, I is the number of insertions and S is the number of substitutions. A deletion is 'a' to '_' (a character to nothing), an insertion is '_' to 'a' (nothing to a character) and a substitution is 'a' to 'b' (a character to another character). Also, a substitution can be viewed as a deletion, followed by a insertion ($a \xrightarrow{D} \_ \xrightarrow{I} b$), that is why the coefficient of S is 2 in the above formula.

In practice, the costs of I, D and S can be changed, but here we use the default costs. Finally, **minimum** edit distance is guaranteed to be found using an algorithm (Damerau–Levenshtein distance) with dynamic programming characteristics.

[3] Levenshtein, V. (1966). "Binary codes capable of correcting deletions, insertions and reversals." Soviet Physice – Doklady 10: 707-710.
[4] Brill, Eric; Moore, Robert C. (2000). An Improved Error Model for Noisy Channel Spelling Correction (PDF). Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. pp. 286–293. doi:10.3115/1075218.1075255. Archived from the original (PDF) on 2012-12-21.