strings and an n-gram representation of each of a set of dissimilar reference character strings, wherein an i-th bit of the binary index key represents a degree of matching of the candidate string with the i-th reference character string.

3. A method in accordance with Claim 2, wherein i-th bit in the binary index key is set to a logical 1 if that record matches the i-th reference string above a predetermined threshold.

4. A method in accordance with Claim 2, wherein the database is managed by a relational database management system (RDBMS) and wherein generating a binary index key comprises forming the binary index key using a process implemented in the RDBMS to use the n-gram frequency similarity calculation that indicates the similarity of a given record to each of the reference strings identified in the PCFA.

5. A method in accordance with Claim 1, further comprising:

computing a two-dimensional vector containing a frequency of occurrence of all unique n-grams in the candidate character string and a frequency of occurrence of all unique n-grams in the reference character string;

computing a similarity metric for the candidate character string, with respect to the reference character string, based on the two-dimensional vector;

determining a magnitude of the vector associated with the candidate character string as magnitude A;

determining a magnitude of the vector associated with the reference character string as magnitude B;

computing a dot product between the two vectors;

computing the similarity metric according to (dot product / (magnitude A × magnitude B)).