Probabilistic String Similarity Joins
Jeffrey Jestes, Feifei Li, Zhepeng Yan, Ke Yi

String-Level Probabilistic qGrams
A string-level probabilistic q-gram is a quadruple $(i, p, s, q)$ where $i$ is the choice index, $(p, s)$ is the begin position, $q$ is the q-gram starting at $i$.

Character-Level Probabilistic qGrams
A character-level probabilistic q-gram is a pair $(\ell, S(\ell..\ell + q - 1))$, where $\ell$ is the beginning position of the q-gram, $S(\ell..\ell + q - 1)$ is the probabilistic substring $S[i] \ldots S[i + q - 1]$.