

Reservoir Sampling: <http://dl.acm.org/citation.cfm?id=3165>

Entity Resolution:

<https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbncndIYmRhdGF8Z3g6MzRjMmFjNzdhM2VIYzlhZA>

https://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance