Similarity Check by Aggregate Profiler

Aggregate Profiler uses “Apache Lucene” core (an open source indexing and searching project - http://lucene.apache.org/java/docs/) to do the similar searches.
Similarity check can be run both on single word and multiword. **Whitespaces ( )** and **Comma (,)** is regarded as delimiter for multiword.
Similarity check can be run on single field or multiple fields. If you choose anything else than “Don’t Use” that field will be used for searching. It is case insensitive search.

Preferred Uses of multiword Search are:
Similar-Any – will match any of the words. Like “Abc Corp. Inc.” will match

- ABCD, international.
- Peculiar corporation ltd.
- New Incorporation

Similar-All – will match all words like “Ramachandran Venugopal” will match

- Venu, Rama
- Venugopaal Ranachandren
- Machandran Venugopalan

“Apache Lucene” uses fuzzy logic to do similarity check.

Preferred Uses of Single Word Search are:
Exact – It will match exact word. If it is multi word, complete phrase will be matched.
Left Imp. – It will match left 4 characters of the word or Phrase
Right Imp. – It will match right 4 characters of the word or Phrase

Importance: Following values will used to boost the importance/relevance of the field in the search criterion. If a field importance is low then even if value is empty or null it will be used in search but if it is medium or high, null/empty value of the field will not used in search.
Low - Default Priority (Boosting value 1)
Medium – Boosting value 2
High – Boosting value 3

Skip Words:
Words put in this text field (comma or space separated) will not be used for search criterion. It is case insensitive. Like – **Inc, Company** is there is in skip box then for Similar-All will match following

- India trading company
- Trading India Inc.

Why Records will appear multiple times: If a record matched Search Criterion, it will not be used for query but if it matches another query also it will appear again. Generally, a subset query will display all it superset records.