Partition keys and corresponding Match keys

|  |  |  |  |
| --- | --- | --- | --- |
| **Product type** | **partition\_key** | **Matching keys(**match\_keys**)** | **Comments** |
| **Journal articles** | **self.partition\_key**=("journal", "difflib", .85) | **self.match\_keys** = ( ((“author”, jarow, 0.85), (“title”, jarow, 0.85)), ) | To be used to match journal articles  "partition\_key":("*journal*", "difflib", .85)  Replace "*journal*" with the name of the field containing the journal name |
| **Books** | **self.partition\_key**= ("year", "identity", 1) | #Partition is year, so don't check again  **self.match\_keys** = ( (( “author”, jarow, .85), (“title”, jarow, .85), (“publisher”, jarow, .6)), ) | To be used to match books  "partition\_key": ("year", "identity", 1)  Replace "*year*" with the name of the field containing the year of publication of the book(in the format yyyy) |
| **Book articles** | **self.partition\_key**=("year", "identity", 1)  #changed from identity to "identity" | #Partition is year, so don't check again  **self.match\_keys** = (  ((“author”, jarow, .95), (“title”, jarow, .85), (“book”, jarow, .85)), ) | To be used to match book articles |
| **Book**  (**amazon\_isbn**) | **self.partition\_key**=("publisher", "difflib", .85) | **self.match\_keys**= (  ((“author”, jarow, .95), (“title”, jarow, .85), (“publoc”, jarow, .6)), ) | To be used to match books :  When year is not provided and place of publication is  "partition\_key": ("publisher", "difflib", .85)  Replace "*publisher*" with the field containing the name of the publisher |
| **Book**  **(ISBNdb\_isbn** ) | **self.partition\_key**= ("publisher", "difflib", .85) | **self.match\_keys** = (  ((“author”, jarow, .95), (“title”, jarow, .85), (“publoc”, jarow, .6), (“year”, identity, 1)),  ) | To be used to match books :  when year and place of publication are provided  Are\*both (**amazon\_isbn**)& **(ISBNdb\_isbn** )  to be run ? When date field present but date data missing in some records???? Answer seems to be yes |

**keys fields values(**use the corresponding fields names in your **table A)**

“year” = year of publication (format yyyy) of book

“author” = author

“title” = journal article title or book title or book article title

“publoc” = place of publication

“book” = title of book

“publisher” = name of the publisher

“journal ” = journal name

**MATCH TABLE**

Create the table that will store the matches with the following command (copy and paste): The table MUST be created in the same database as table A

you may change the table name but not the field names.

this (table name)will be your table\_matches attibute in the child class

CREATE TABLE `relships\_sameas` (`relid` mediumint(8) unsigned NOT NULL auto\_increment,

`id1` varchar(30) NOT NULL default '0',`table1` varchar(64) NOT NULL,

`id2` varchar(30) NOT NULL default '0',`table2` varchar(64) NOT NULL,

`confidence` int(3) unsigned NOT NULL default '0',PRIMARY KEY (`relid`),

KEY `id2` (`id2`),KEY `id1` (`id1`),KEY `table1` (`table1`),

KEY `table2` (`table2`)\n) ENGINE=MyISAM DEFAULT CHARSET=latin1