Match References to Papers

Menu Path

Data Preparation > Database > ISI > Match References to Papers

Description

The use of the word "Paper" and "References" in this algorithm name is a misnomer. It should actually be *Match Citations to Documents* to maintain technicality.

This algorithm attempts to match Citations to Documents (in an "is a" relationship). A citation is considered to match a document if and only if:

- the Citation Author, Page Number, Source, Volume, and Year are all provided and are valid;
- the Citation Author matches the first Author of the document ;
- the Citation Page Number matches the document Beginning Page;
- the Citation Source and Document Source are exactly the same Source;
- the Citation Volume matches the document Volume;
- the citation Year matches the document Year.

The output of this algorithm is a copy of the input database, but with the Citations Table table updated to point to the Documents Table table (via the *document_id* field). When it is finished, it reports how many citations were matched to documents and how many citations were over-matched to documents. An over-matched citation is a citation that matches more than one document.

Usage Hints

Load an ISI file into the tool, then create a database from it using the ISI database loader.

It is strongly recommended that the database be cleaned before matching citations to documents.

For a quick analysis of a small dataset you may wish to merge together author entities with identical names. For a scientifically sound analysis of a larger dataset, you can find author entity merging suggestions (or manually set your own merging orders from scratch) and perform the merge.

Then, you will probably want to merge together journal entities according to recognized variants.

Links

- Source code
- The string templates of the database queries used to determine the matches and over matches from References to Documents.
- Load ISI File into Database
- Merge Identical ISI People
- Merge Document Sources