Extract Longitudinal Summary

Menu path

Data Preparation > Database > ISI > Longitudinal Summary

Description

Extracts a longitudinal summary from an ISI database.

Typically, each document in your dataset will have been published in a particular year and each reference for each document in your dataset will also be to some document published in a particular year.

This algorithm produces a table which contains one row for each of those years and provides counts of a variety of entities and events from that year.

Analyses

The output table will include the following summaries of your dataset for each publication year and each referenced year:

- *documents_published*: The number of documents published that year.
- *references_published*: The number of documents which refer to some document published that year.
- *total_references_made*: The total number of references (and *distinct_references_made*: the number of distinct references) cited among all of the documents published that year.
- *distinct_authors*: The number of distinct authors who published a document that year.
- *distinct_sources*: The number of distinct sources (journals, typically) that contain a document published that year.
- *distinct_author_keywords*: The number of distinct author-provided keywords among all documents published that year.
- *distinct_isi_keywords*: The number of distinct ISI-provided keywords among all documents published that year.
- *distinct_other_keywords*: The number of distinct keywords among all documents published that year that were not provided by the author(s) or by ISI.

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 extracting the longitudinal summary.

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.

Finally, you must match references up to documents in your dataset.

Implementation Details

The specific query run by the tool can be found in the source code.