Load ISI File Into Database

Description

Accepts ISI file (of the human-readable flavor) as the input and produces an ISI database as the output. This is the only provided solution for loading an ISI file into a database.

As with our other database loaders, loading happens in three phases. In the first phase, the data is loaded from disk and processed so it can then be inserted into a database. In the second phase, a new database is created and the loaded data is inserted into it. In the third phase, the document sources are automatically merged using Merge Document Sources.

The progress Phase 1 is tracked in the "% Complete" column in the Scheduler. When it finishes, the "% Complete" is reset for Phase 2, which also tracks its progress there. Currently the progress monitor is not working for phase three.

Upon the beginning of the first phase, the message "Beginning Phase 1 of loading the ISI file into a database." is printed to the Console. Upon the beginning of the second phase, the message "Beginning Phase 2 of loading the ISI file into a database." is printed to the Console.

Usage Hints

Load a human-readable ISI file (so not a tabular/CSV one), then run this algorithm on it. This must be done as the first (set of) step(s) in the database pipeline.

Note: ISI does not provide a conclusive list of fields (or "tags" as we often call them) supported in neither their *Web of Knowledge* nor *Web of Science* file formats. However, it appears to be the case that there are fields that *Web of Knowledge* datasets include that *Web of Science* datasts do not include. The database schema we use in this ISI loader and our other ISI database-related algorithms takes into account only the most commonly-used *Web of Science* fields. All ISI files are supported, but any "arbitrary" fields found will merely be appended to the **DOCUMENT** table and will not be reflected in the schema structure on a high level.

Implementation Details

This loader first converts the ISI data from the human-readable format to a table. It then transforms that table to a database. No data is lost or changed during either of these steps, but some linkages/metadata are created where possible.

The following is a list of ISI database tables that link to pages describing their fields and how they are parsed out of ISI datasets:

- ADDRESS
- AUTHORS
- CITED_PATENTS
- CITED_REFERENCES
- DOCUMENT
- DOCUMENT_KEYWORDS
- DOCUMENT_OCCURRENCES
- EDITORS
- ISI_FILES
- KEYWORD
- PATENT
- PERSON
- PUBLISHER
- PUBLISHER_ADDRESSES
- REFERENCE
- REPRINT_ADDRESSES
- RESEARCH_ADDRESSES
- SOURCE