

# DOCUMENT

This table contains data directly pertaining to documents. It includes all fields relevant for all possible (known) types of documents in ISI *Web of Science* data; that is, a Document is not necessarily always a(n) article, conference, paper, etc.

All "arbitrary" fields found in the original ISI dataset are also appended to this table, as this table is somewhat central to the database schema. Any fields that are not in the following list are considered to be arbitrary:

- AB (Abstract)
- AR (New Article Number)
- AU (Authors)
- AF (Author Full Names)
- BA (Book Authors)
- BE (Edited By)
- BF (Book Author Full Name)
- BP (Beginning Page)
- SE (Book Series Title)
- BS (Book Series Subtitle)
- CP (Cited Patent)
- NR (Cited Reference Count)
- CR (Cited References)
- CY (Cited Year)
- OI (ORCID ID)
- RI (Author Identifier)
- HO (Conference Host)
- CL (Conference Location)
- SP (Conference Sponsors)
- CT (Conference Title)
- MA (Meeting Abstract)
- DT (Document Type)
- DI (DOI)
- D2 (Book DOI)
- ED (Editors)
- EM (E-mail Addresses)
- EF (End of File)
- EP (Ending Page)
- ER (End of Record)
- FN (File Type)
- SO (Full Journal Title)
- FU (Funding Agency and Grant Number)
- FX (Funding Text)
- BN (ISBN)
- SN (ISSN)
- GA (ISI Document Delivery Number)
- JI (Abbreviated ISO Journal Name)
- IS (Issue)
- LA (Language)
- ID (New ISI Keywords)
- PG (Number of Pages)
- DE (Original Keywords)
- PN (Part Number)
- PD (Publication Date)
- PT (Publication Type)
- PY (Publication Year)
- PA (Publisher Address)
- PI (City of Publisher)
- PU (Publisher)
- WP (Publisher Web Address)
- RP (Reprint Address)
- C1 (Research Addresses)
- SI (Special Issue)
- SC (Subject Category)
- SU (Supplement)
- TC (Times Cited)
- TI (Title)
- J9 (Abbreviated Journal Name)
- UT (Unique ID)
- VR (Version Number)
- VL (Volume)

Note that the above is merely a list of ISI fields handled by the loader, not fields that end up in this table. That being said, this table contains the following fields:

- *PK*: Automatically generated. Guarantees uniqueness.
- *ABSTRACT\_TEXT*: The exact string as found in the "AB" field.
- *ARTICLE\_NUMBER*: The exact string as found in the "UT" field.
- *BEGINNING\_PAGE*: The contents of the "BP" field parsed as an integer.
- *CITED\_REFERENCE\_COUNT*: The contents of the "NR" field parsed as an integer.

- *DIGITAL\_OBJECT\_IDENTIFIER*: The exact string as found in the "DI" field.
- *DOCUMENT\_TYPE*: The exact string as found in the "DT" field.
- *DOCUMENT\_VOLUME*: The contents of the "VL" field parsed as an integer.
- *ENDING\_PAGE*: The contents of the "EP" field parsed as an integer.
- *FIRST\_AUTHOR\_FK*: If any authors were provided (in the "AU" field), this will be the foreign key into the **PERSON** table of the first author specified. See [AUTHORS](#) and [PERSON](#).
- *FUNDING\_AGENCY\_AND\_GRANT\_NUMBER*: The exact string as found in the "FU" field.
- *FUNDING\_TEXT*: The exact string as found in the "FX" field.
- *ISBN*: The exact string as found in the "BN" field.
- *ISI\_DOCUMENT\_DELIVERY\_NUMBER*: The exact string as found in the "GA" field.
- *ISI\_UNIQUE\_ARTICLE\_IDENTIFIER*: The exact string as found in the "UT" field.
- *ISSUE*: The exact string as found in the "IS" field.
- *LANGUAGE*: The exact string as found in the "LA" field.
- *PAGE\_COUNT*: The contents of the "PG" field parsed as an integer.
- *PART\_NUMBER*: The exact string as found in the "PN" field.
- *PUBLICATION\_DATE*: The exact string as found in the "PD" field. Is not an actual date field in terms of data types because it may actually be a range or list of dates, or it may contain arbitrary text.
- *PUBLICATION\_YEAR*: The contents of the "PT" field parsed as an integer.
- *DOCUMENT\_SOURCE\_FK*: If a source was provided (see [SOURCE](#) for an explanation of which ISI fields end up in the **SOURCE** table), this will be the foreign key into the **SOURCE** table.
- *SPECIAL\_ISSUE*: The exact string as found in the "SI" field.
- *SUBJECT\_CATEGORY*: The exact string as found in the "SC" field.
- *SUPPLEMENT*: The exact string as found in the "SU" field.
- *TIMES\_CITED*: The contents of the "TC" field parsed as an integer.
- *TITLE*: The exact string as found in the "TI" field.

It is possible for References to be matched to Documents (as an "is a" relationship) during loading based on either Digital Object Identifiers and/or Article Numbers matching. See [REFERENCE](#) for more details on this. (There is also an explicit [Match References To Papers](#) step, which uses different metrics for matching Documents to References.)

In the various built-in extractions, when Documents are specified in the original ISI dataset, they are considered to be inner documents. When they are only specified as References and are not matched to actual Documents, they are considered to be outer documents.