
The Cystic Fibrosis Database (CF) consists of 1239 documents published from 1974 to 1979 discussing Cystic Fibrosis Aspects, and a set of 100 queries with the respective relevant documents as answers.

The original collection is available in a single [gzipped tar file](#) or a [zip file](#) both of 1.47Mb, containing 7 document files and 1 query file.

The collection is also available in XML format also in a single [gzipped tar file](#) or a [zip file](#) both of 1.54Mb, including the Document Type Definition (DTD) for the collection and for each query/answer.

Document Files

Each document includes 11 fields as follows:

Paper Number The first two digits give the year of publication, and the rest three digits range from 1 to the number of docs published that year. **Record Number** serial id number varying from 1 to 1,239. **Medline Accession Number** CF is a subset of the MEDLINE database. **Author(s)** **Title Source** Bibliographic citation of source. **Major Subjects** The Medical Subject Headings (MeSH) and subheadings representing the major subjects of the document. The Medical Subject Headings are shown in capital letters and have been assigned by expert indexers. The two-letter symbols are subject subheadings, also assigned manually from a controlled vocabulary (see the MeSH vocabulary published by the National Library of Medicine). **Minor Subjects** The Medical Subject Headings (MeSH) and subheadings representing the minor subjects of the document. The Medical Subject Headings are shown in capital letters and have been assigned by expert indexers. The two-letter symbols are subject subheadings, also assigned manually from a controlled vocabulary (see the MeSH vocabulary published by the National Library of Medicine). **Abstract/Extract** The abstract of the document, or in the case of a document with no published abstract, an extract from text. **References** The complete list of references appearing in the document, excluding private communications and unpublished documents. **Citations** A comprehensive list of citations to the document, as indexed in the SCISEARCH/DIALOG files

Query Files

Each query includes a query number and text, the record number of each relevant document in the answer, and relevance scores.

The relevance scores are from 4 different sources: REW (one of the authors), faculty colleagues of REW, post-doctorate associate of REW, and JBW (other author and a medical bibliographer).

The relevance scores vary from 0 to 2 with the following meaning:

- 2 HIGHLY relevant
- 1 MARGINALLY relevant
- 0 NOT relevant

Example of a document answer: 513 0010

Doc number: 513

Relevance score by REW: NOT relevant.

Relevance score by REW colleagues: NOT relevant.

Relevance score by REW post-doctorates: MARGINALLY relevant

Relevance score by JBW: NOT relevant.

Copyright

This collection is available thanks to the original authors from the School of Information and Library Science, University of North Carolina, Chapel Hill, NC 27599-3360, USA.

They have the copyright (1989) and the reference to their work is:

- Shaw, W.M. & Wood, J.B. & Wood, R.E. & Tibbo, H.R. The Cystic Fibrosis Database: Content and Research Opportunities. LISR 13, pp. 347-366, 1991.

The citations in the CF document collection represent a small subset of MEDLINE data and should not be used to search for current references on the subject of cystic fibrosis.