

**005.7, 005.74 vs. 025.04, 025.06****Data sets, data files, databases**

Data sets are collections of data, possibly unstructured or variously structured. Data files and databases, although there are technical differences between them, are treated as the same for classification purposes and are a specific kind of data set.

Use 005.7 for the algorithms used in storing and retrieving collections of data. Use 005.74 for designing and searching databases, for programming and installing database management systems, in which the databases contain data that is structured or relatively structured.

Use 025.04 or 025.06 for indexing and searching bibliographic and relatively unstructured text databases.

Use 001–999 for the subject content of data sets and databases (and works discussing that content) as if the data sets and databases were books, e.g., chemistry data sets or databases 540. Do not use notation 028557 or notation 0285574 from Table 1 except for works that focus on the computer science aspects of the data sets or databases rather than the subject content.

If in doubt, prefer 005.7 and 005.74.

*See also discussion at 011.39 vs. 005.3029, 016.0053, 025.0422.*