Databases

Definition

Any collection of data, or information, that is organized for rapid search and retrieval by a computer.

Organization / Structure

Databases are structured to facilitate the storage, retrieval, modification, and deletion of data in conjunction with various data-processing operations.

Bibliographic databases are collections of descriptions of published information. Usually the description includes the citation (author, title, source) and a summary or abstract. Occasionally the full text of the described publication is included or a link to the publication may be provided if the publication is available on the web.

- A **bibliographic database** usually consists of a **file** or a set of **files**.
  - **Files** consist of **records**.
    - **Records** consist of one or more **fields**

Fields are the basic units of data storage, and each field typically contains information pertaining to **one aspect or attribute of the entity** described by the **record**.

Source

"Database" *Encyclopædia Britannica* from Encyclopædia Britannica Online . http://search.eb.com/eb/article?tocId=9029424 [Accessed January 7 2005].

Example of **fields** in a **record**

This Medline record illustrates how a bibliographic record is divided into fields:

Medline Record	
Field Label	**Field Contents**
TI:	A quantitative assessment of ancient therapeutics: poppy and pain in the Hippocratic Corpus.
AU:	Prioreschi-P; Heaney-RP; Brehm-E
AD:	Department of Pharmacology, Creighton University, Omaha, NE 68178-0001, USA.
SO:	Med-Hypotheses. 1998 Oct; 51(4): 325-31
ISSN:	0306-9877
PY:	1998
LA:	ENGLISH

AB:	In an effort to develop a method to ascertain whether ancient physicians used certain drugs because of their efficacy or for other reasons (random process, tradition, etc.), the authors have devised a statistical test (Efficacy Quotient-EQ) to evaluate ancient therapeutic data quantitatively. They have applied the Efficacy Quotient to clarify whether the Hippocratic physician used poppy (opium) in the treatment of pain because he had recognized that it had analgesic properties or whether he used it as he used others plants devoid of such properties. The authors conclude that the analgesic properties of poppy were not recognized at the time. The appropriateness of using the Efficacy Quotient for the.
MESH: (minor)	History-of-Medicine,-Ancient; Opium-therapeutic-use; Pain-drug-therapy; Papaver-
MESH: (major)	*Opium-history; *Pain-history; *Statistics-
TG:	Female; Human; Male
PT:	HISTORICAL-ARTICLE; JOURNAL-ARTICLE
RN:	8008-60-4
NM:	Opium
AN:	1999042281

Search Mechanics

A typical search looks like this:

Set	Records	Search Statement
#1	3	Consumption AND Fruit

This search retrieves all the records in the database containing both the words *consumption* and *fruit*. The records retrieved by the search statement are held in a set (#1 in this example).

The same search could be entered term by term and the resulting sets combined in the last step. The result is the same:

Set	Records	Search Statement
#1	31530	Consumption
#2	3597	Fruit
#3	3	#1 AND #2

This second search provides a better illustration of the mechanics of a search, although the process is the same in both searches. When the computer conducted the first search, *Consumption AND Fruit*, it temporarily created the equivalent of sets #1 and #2 in the second example. The only difference is that in the first search these sets were never displayed and were erased when the search was complete.

Question:

What happens when you ask the computer to find all the records containing both the words *fruit* and *consumption* - does the computer actually scan each record to see if both words are present?

Answer:

No, the computer never actually searches the records themselves. To find out what actually happens read on ...

Recall that:

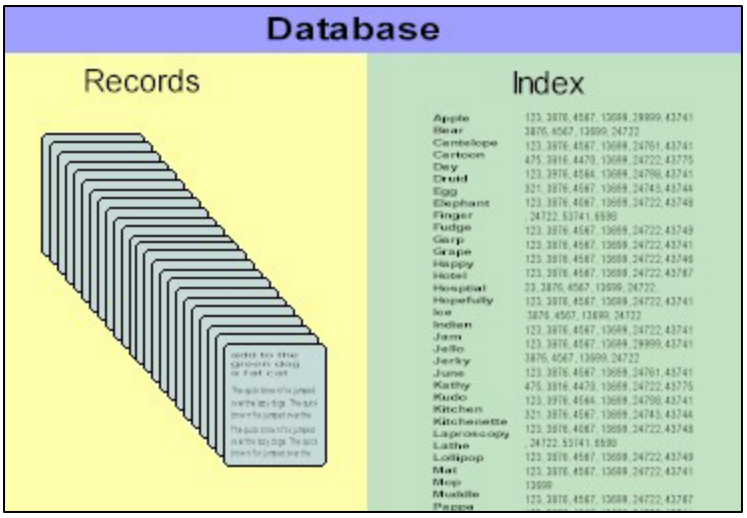
Databases are made up of **files**,
and **files** are made up of **records**,
and **records** are made up of **fields**.

Bibliographic databases have at least two **files**:

- A file of records
- An index file.

The index file contains an alphabetic list of every *"word"* that occurs in the records.

Associated with each word in the index is a list of every record number in which that word occurs. Every record in the database is assigned a unique record number.



Excerpt: Alphabetical index of all terms in the database.

Word	No. of Records	Record Numbers
:		:
CONSUMED	5779	54, 256, 311, 327, 1625, 2412, 2841, 4137, 6910, 7102, 7184, 7509, ...
CONSUMER	5340	245, 258, 275, 2458, 2572, 5285, 5475, 5878, 7521, 7892, 8587, ...
CONSUMERS	15340	68, 655, 864, 978, 5358, 6546, 7981, 8732, 9841, 9872, 10897, ...
CONSUMES	194	237, 394, 679, 1816, 2398, 4981, 5791, 7910, 8704, 8741, 9487, ...
CONSUMING	4696	91, 384, 854, 1583, 2896, 4579, 6587, 7845, 9654, 9874, 10952, ...
CONSUMMATE	13	586, 853, 1102, 1479, 2487, 3256, 3489, 5749, 6187, 7421, 8271, ...
CONSUMMATED	4	478, 3062, 7903, 8385
CONSUMMATES	1	9523
CONSUMMATING	3	3076, 5341, , 10204
CONSUMMATION	9	2087, 3956, 4592, 4781, 6134, 6136, 7492, 8129, 8476, ...
CONSUMMATORY	138	1063, 1547, 2079, 2492, 3256, 3533, 3947, 7845, 9248, 10245, ...
CONSUMPTION	31530	70, 256, 311, **467**, **829**, 1625, 2841, 3527, 4173, 4431, 4918, 5081, ...
:		:
FRUGALITY	13	1479, 1927, 3890, 4681, 5578, 5988, 6240, 8134, 8348, 8639, 98233, ...
FRUGALLY	4	2074, 5729, 6038, 7365
FRUGIPERDA	618	494, 1682, 1928, 2045, 2321, 5566, 6289, 6430, 7740, 7860, 8252, ...
FRUIT	3597	54, 256, **467**, **829**, 898, 2412, 4137, 4173, 5081, 6041, 7959, 8166, ...
FRUITBATS	5	223, 24130, 4216, 7792, 9956
FRUITCAKE	1	1045
FRUITED	10	582, 725, 1698, 1938, 2580, 3340, 3671, 4001, 4248, 6812
FRUITFLIES	24	2597, 3019, 3620, 4124, 4550, 5396, 5612, 5850, 6070, 64329, 7381, ...
FRUITFLY	77	555, 1020, 2234, 2597, 3500, 4124, 5072, 5396, 5941, 5942, 7709, ...
FRUITFULLY	34	307, 905, 4583, 4867, 5280, 6007, 6432, 6667, 7123, 8765, 8839, ...
FRUITFULNESS	9	35, 3166, 4809, 4952, 5588, 5712, 6034, 8842, 9631
FRUITINESS	3	4059, 6398, 8954
FRUITING	317	739, 1139, 1972, 2082, 2531, 3063, 3544, 5671, 6137, 7028, 7030, ...
FRUITLESS	45	1089, 2946, 3103, 3905, 4093, 5623, 5932, 6410, 6749, 8678, 10309, ...
FRUITS	1644	54, 70, 226, 311, 836, 1625, 2412, 4173, 5414, 6041, 8329, 9207, ...
FRUITY	15	4038, 4039, 4040, 6398, 7485, 8302, 8623, 8678, 92018, 10245, ...
:		:

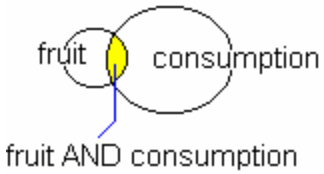
When a search for all records containing both the words FRUIT AND CONSUMPTION is submitted the following steps occur:

1. The computer looks up *fruit* in the index.
2. When it finds *fruit* it retrieves all the associated record numbers and holds them as a set.
3. The computer then looks up *consumption* in the index.
4. When it finds *consumption* it retrieves all the associated record numbers and holds them as a set.
5. It compares the record numbers in the two sets.
6. Any record number which occurs in both sets is a *hit* and that number is put into a third set - the retrieval set.
7. Watch this *illustration* of how two sets of numbers are processed in an 'AND' operation.
8. When the records are displayed from the retrieval set, the records are fetched by record number from the database's records file.

This chart shows the sets and record numbers for this process:

Set	Word	Hits	Record Numbers
A	**CONSUMPTION**	31530	70, 256, 311, **467**, **829**, 1625, 2841, 3527, 4173, 4431, 4918, 5081, ...
B	**FRUIT**	3597	54, 256, **467**, **829**, 898, 2412, 4137, 4173, 5081, 6041, 7959, 8166, ...
C	**A** AND **B**	5	256, 467,829,4173, 5081

This Venn diagram shows the outcome of the search in yellow:



Now that you understand the mechanics of searching, you should see that Venn diagrams provide a simple but accurate picture of the outcome of a search.

This is one of the records retrieved by the "*fruit and consumption*" search:

RN:	**256**
TI:	Fruit and vegetable consumption in later life.
SO:	Age-Ageing. 1998 Nov; 27(6): 723-8

This records was **not** retrieved because it doesn't contain the word *fruit* (singular). If you look at the sample index again, you will see that record number 311 is in the list for both *consumption* and *fruits* (plural), but it is not in the list for fruit.

RN	**311**
TI:	Factors affecting consumption of fruits and vegetables by low-income families.
SO:	J-Am-Diet-Assoc. 1994 Nov; 94(11): 1309-11

Moral:
What the computer does is very mechanical. You have to provide the strategy for a successful search. If you keep in mind that a successful search is *"just"* a matter of finding the right term combinations, you are on your way to becoming a good searcher.