

Langual for Database Users

Elizabeth C. Smith, M.S.L.S.
Food and Drug Administration
Washington, District of Columbia

Finding nutritional, consumption or toxicological information on food can be as difficult as finding one's way through a rat maze. The Center for Food Safety and Applied Nutrition (CFSAN) of the U.S. Food and Drug Administration (FDA) has developed an excellent information retrieval system that links dissimilar databases through a common descriptive language. The information retrieval system is known as the **Food Monitoring Database** and the descriptive system is called **Langual**. The name comes from French words meaning "Universal Language for Food." Langual has been presented at previous National Nutrient Databank Conferences. For a detailed description of its construction and features, see page 93 of the Proceedings of the Fourteenth National Nutrient Databank Conference, June 19-21, 1989, University of Iowa, Iowa City, Iowa.

Langual is a standardized vocabulary for food product description. It is composed of fourteen different viewpoints or factors such as "Product Type" or "Extent of Heat Treatment." The factors contains numerous descriptive terms, each of which may be used to retrieve information about a food product. Langual provides assistance to users in a number of ways. Definitions explain what a term is or how it is meant to be used. Synonyms for scientific nomenclature or vernacular usage are available; French synonyms were added in the past year. Retrieval terms are arranged in a hierarchy, which arrays terms conceptually from broader to narrower.

Langual is used for information retrieval in CFSAN's Food Monitoring Database (FMDB). It is stored at the Parklawn Computer Center in Rockville, Maryland, on an IBM 3090 K64, using Model 204 Database Management System. Nine diverse food databases have been indexed by using Langual. Six of these files

are from sources outside the FDA. They are the USDA Nutrient Database for Standardized Reference (Handbook 8); the 1987/88 Nationwide Food Consumption Survey; food names from the Codex Alimentarius; a carotenoid food file; a French food file and a Greek food file. The three remaining food files are FDA-based. They are the Total Diet Study (TDS); the Food Labeling and Product Surveillance (FLAPS); and the Scientific Information Retrieval and Exchange Network (SIREN). More than 21,000 food products are indexed by Langual and searchable in the FMDB.

Nutritionists, food scientists and toxicologists have different information needs. Such questions as **What is the iron content of sugared breakfast flakes?** or **Do sugared breakfast flakes contain malathion?** may be answered by using Langual. Both questions may be answered by finding all **BREAKFAST CEREALS** that are **FLAKED** and **CONTAIN SUGAR**. Each of these elements is a retrieval term in Langual. Definitions explain that the "contains sugar" element is searched by using the index term **SUCROSE ADDED**.

When **BREAKFAST CEREAL** is entered into the Food Monitoring Database, food names from almost all the databases are retrieved. However, when the index terms **FLAKED** and **SUCROSE ADDED** are combined with the retrieval from **BREAKFAST CEREAL**, a much smaller number of food names is retrieved.

Example:

# Food Names	Database
29	Nationwide Food Consumption Survey
4	SIREN
2	Total Diet Study
20	USDA Handbook 8
3	FLAPS
8	Carotenoid Foods

The names of the food products are then displayed. Each database is listed separately and each food name has a code identifier linking it to the database.

Example:

USDA Handbook 8

Corn flakes cereal, low sodium	FOODID 08022
Team cereal	FOODID 08075

Total Diet Study

Cereal, cornflakes	TDCODE 071
Cereal, raisin bran	TDCODE 074

In summary, Languag not only provides a way of retrieving foods with common characteristics from a single database, but it also provides the capability for linking highly dissimilar databases through a universal retrieval language.

When food products sharing common characteristics are identified, the Food Monitoring Database search program is used to find nutritional or toxicological data. USDA Handbook 8, which contains results from nutritional analyses of foods, is used to answer the first question on the iron content of sugared breakfast flakes. The word **IRON** is entered into the FMDB and arrayed in an alphabetical list. It is then selected for retrieval and matched with the breakfast cereal food names found in Handbook 8. From the USDA nutritional data stored in the FMDB, a final report displays the results.

Example:

USDA HANDBOOK #8 REFERENCES

<u>ID #</u>	<u>Description</u>	<u>Mean # Obs.</u>	<u>Std. /100 g</u>	<u>Error</u>
08022	Corn flakes cereal, low sodium IRON	2	2.22	.177
08070	Team cereal IRON	21	6.12	.176

For the question on malathion the Total Diet Study would provide the answer. The word **MALATHION** is entered into FMDB and arrayed alphabetically. It is then selected and matched with the breakfast cereal names previously retrieved. A final report is displayed showing residue amounts of malathion in parts per million using TDS toxicological data.

Example:

TOTAL DIET STUDY REFERENCES

<u>TDCODE</u>	<u>Description</u>	<u>Basket Year</u>	<u>Residue Amount</u>	<u>Residue Unit</u>
071	Cereal, cornflakes	1981	0.002	ppm
074	Cereal, raisin bran	1985	0.700 0.006 0.015	ppm
		1986	0.004 0.015	ppm