

UPDATE ON THE FACTORED FOOD VOCABULARY (LANGUAL)

Elizabeth C. Smith, M.S.L.S.

*Food and Drug Administration
Center for Food Safety and Applied Nutrition
Washington, D.C.*

The Factored Food Vocabulary is a computer-based information retrieval tool developed at the Center for Food Safety and Nutrition (CFSAN) of the Food and Drug Administration (FDA). Work began on it in 1975 and continues today by a group known as the LANGUAL Developmental Working Committee. During the past year the Factored Food Vocabulary was renamed LANGUAL and is now being viewed as an international language for foods. LANGUAL provides a standard vocabulary for the description of food products in databases. Its structure is based on the idea that a food product may be described by a combination of several characteristics, each of which may serve as a point of retrieval. LANGUAL is composed of 13 viewpoints or "factors":

PRODUCT TYPE
FOOD SOURCE
PART OF PLANT OR ANIMAL
PHYSICAL STATE, SHAPE OR FORM
EXTENT OF HEAT TREATMENT
COOKING METHOD
TREATMENT APPLIED
PRESERVATION METHOD
PACKING MEDIUM
CONTAINER OR WRAPPING
FOOD CONTACT SURFACE
CONSUMER GROUP/DIETARY USE
ADJUNCT CHARACTERISTICS OF FOOD

Each factor contains a number of factor terms or descriptors, one or several of which may be selected to describe a food product. With most factors, a single descriptor is selected; however, multiple factor terms may be employed in TREATMENT APPLIED, COOKING METHOD, FOOD CONTACT SURFACE, and for special dietary characteristics found in CONSUMER GROUP/DIETARY USE. A food product such as corn flakes is described as follows:

FACTOR

PRODUCT TYPE
FOOD SOURCE
PART OF PLANT OR ANIMAL

PHYSICAL STATE, SHAPE OR FORM

DEGREE OF PREPARATION
COOKING METHOD
TREATMENT APPLIED

PRESERVATION METHOD
PACKING MEDIUM
CONTAINER OR WRAPPING
FOOD CONTACT SURFACE
CONSUMER GROUP/DIETARY USE

FACTOR TERM

Breakfast cereal
Field Corn
Seed or kernel, skin removed, germ removed (endosperm)
Whole, shape achieved by forming, thickness <0.3 cm.
Fully cooked
Cooking method not applicable
Sucrose added; Flavoring or spice extract or concentrate added; Vitamin added; Iron added; Flaked; Water removed
Dehydrated or dried
No packing medium used
Paperboard container with paper liner
Wax; Paperboard or paper
Human food, no age specification, regular diet

Many factor terms in LANGUAL have definitions describing what they are, what they include or exclude, and how they are intended to be used within the vocabulary.

Brown sugar

Food dictionary -- soft sugar whose crystals are covered by a film of refined dark syrup that imparts color, flavor and moisture.

LANGUAL also contains many synonyms, including scientific names for organisms and vernacular terms.

Citrus sinensis	USE	Orange
Shredded	USE	Divided into pieces, thickness <0.3 cm.

Finally, each set of terms for a specific factor is arranged hierarchically from broader to narrower terms. This arrangement provides a meaningful classification that relates terms within a factor to each other.

Carbohydrate

Sugar

Fructose

Lactose

Sucrose

Brown sugar

White sugar

At CFSAN, an information retrieval system called the Food Monitoring Database uses LANGUAL to link food names in three databases: FDA's Total Diet Study, which monitors levels of contaminants and nutritional elements in the U.S. food supply; the U.S. Department of Agriculture (USDA) Nutrient Database for Standard Reference, which contains nutrient analytical data; and FDA's Scientific Information and Retrieval Exchange Network (SIREN), which contains bibliographic citations of FDA regulatory and petition information. To search the Food Monitoring Database, LANGUAL factor terms describing food products are entered into the program; food product names that meet the description criteria are simultaneously retrieved from each database. Once a list of names is retrieved, each database may be queried individually for analytical or bibliographic data. For example, the SIREN file may be searched for regulatory or petition citations to orange juice, the USDA Handbook No. 8 file might be searched for mean levels of ascorbic acid in orange juice, or the Total Diet Study may be searched for contaminant levels of carbaryl in orange juice. Thus, LANGUAL provides a common language for linking dissimilar files.

During the past year efforts at CFSAN have been directed toward indexing a backlog of food product names from SIREN. Since last September more than 2,000 foods have been indexed; approximately 2,500 remain to be indexed. When this long-range effort is complete, approximately 7,000 food product names will have been indexed. This volume of indexing has led to software enhancements designed to facilitate the indexing process. The most important of these enhancements is automatic factoring or "Autofactor", with which new database with new food names is entered in the Food Monitoring Database, each new food name is compared with existing food names in the file. If a match or a high degree of similarity is found, LANGUAL factor terms from the existing food name may be automatically selected to describe the new food. Additional factor terms may be added to complete the food description. This technique has substantially reduced the time needed to index a food product and has ensured consistency of indexing. For those who might consider using LANGUAL, Autofactor is a tool that should not be overlooked.

Another software enhancement, which is still in the developmental phase, is the display of a portion of hierarchy and a definition within the indexing context. Formerly, hierarchy, definitions, and factor terms were viewed in one file and then index terms were applied from another. Now both these functions are available in a single computer program. Since one of the basic rules for LANGUAL is to index as specifically as possible, this enhancement provides a way to check that the most specific factor term has been used.

Additionally, once the portion of hierarchy is displayed, the definition for the displayed term may be viewed, ensuring that the term is used correctly.

The LANGUAL vocabulary continues to grow, with new terminology being introduced from SIREN and international indexing efforts. Requests from France, Denmark, and England have been reviewed by the LANGUAL Developmental Working Committee. The committee is devising a factor that arrays scientific names of food sources by taxonomic classification and is considering methods of expanding food product descriptions to include all ingredients.

New databases will be indexed by LANGUAL and added to the Food Monitoring Database. The food composition database used for the 1987-88 Nationwide Food Consumption Survey and for the National Health and Nutrition Examination Survey (NHANES III) as well as CFSAN's Food Labeling and Product Surveillance (FLAPS) files are under consideration. International interest and usage is expected to continue. Thus, LANGUAL will continue to provide a unifying language to describe food products and a means of linking databases dissimilar in origin, structure, or use.