A language for the description of foods in databases

Dagobert Soergel
College of Information Studies, University of Maryland

Anne McCann
Division of Information Resources Management
Center for Food Safety and Applied Nutrition
Food and Drug Administration

Joanne Holden
Nutrient Composition Laboratory
Human Nutrition Research Center
Beltsville Agricultural Research Service
USDA

Jean Pennington
Division of Nutrition
Center for Food Safety and Applied Nutrition
Food and Drug Administration

Abstract

This paper presents a highly flexible and powerful language for the description of foods in food composition and food use databases and for the precise specification of recipes and food standards. The language is the conceptual foundation for a proposed master database of food descriptions which would offer vastly improved capabilities for retrieval and combination of data from food composition and food use databases. The language incorporates principles from classification theory, database theory, and artificial intelligence. It uses the **entity-relationship approach** to structuring food data in a way that mirrors the structure of the foods themselves. The entity types and relationship types needed to build food descriptions are detailed in the **conceptual schema**, and the most important ones are arrayed in a **frame** to guide indexers and searchers. The data structure uses **hierarchical inheritance** to avoid redundancy in storage. This paper also discusses **food names and codes** and their relationship to food description. It defines **explicit names or codes** that carry their own explanation and brief **nonexplicit names or codes** that merely refer to an explanation stored in one's memory or a database. In this context the paper discusses **levels of specificity of the food products named**, from product type to individual sample analyzed or portion consumed.

Table of Contents

1   Introduction

This paper presents a highly flexible and powerful language for the description of foods. The language is designed for use with an interlinked database structure and is therefore called Interlinked Food Description Language. It is designed for the description of foods in food composition and food use databases and for the precise formal specification of recipes and food standards. Development of the language was motivated by the insight that lack of detail and standardization in food description is a major barrier to the retrieval of food data and to the combination of data from several databases (see the first paper in this two-part series). To lessen this barrier, we proposed a master database of complete and detailed standardized food descriptions, which would be linked to many other databases that contain food composition, food use, and other food-related data. Thus the master database could serve as a convenient central switching point for requesting data from one or more of the many food and nutrition databases. Such a master database requires as its conceptual foundation a standardized language for the description of foods.

Another important issue in the coordination of food databases is compatibility and standardization of food names and codes. Food naming and food description are closely related, and consistent codes for the identification of foods play a central role in database structure. Therefore, we discuss first food names and codes, their relationship to food description, and their varying levels of specificity as a preliminary to the main topic, the structure of a language for the description of foods and the associated interlinked database structure.

This paper grew out of the work on the Factored Food Vocabulary (FFV, now called Langual for Langue Alimentaire) developed since 1975 at the Food and Drug Administration, Center for Food Safety and Applied Nutrition (FDA/CFSAN) by the FFV Committee, an interdisciplinary group with expertise in nutrition, food technology, and information/computer science (1) (2) (3). This system uses the same principle, facet analysis, as the International Feed Databank System (4) (5) (3). The further developments proposed here take advantage of the advances in artificial intelligence, database management, and programming languages since work on the FFV began.

2   Food names and food descriptions

Foods are identified by names or codes so that they can be referred to in product labels, survey instruments, documents, handbooks, files, and databases. Often a food name describes a food more or less completely, for example, *sugar-coated puffed wheat* or *spicy chilled tomato soup*. Thus the issue of designating foods by names or codes is related to the issue of describing foods. (However, food designations do not have to convey descriptive information.)

This section examines the intertwined issues of food naming and food description. Table 1 summarizes the new terms introduced.

2.1 Explicit and nonexplicit food names

Many food names consist of component terms, each referring to an aspect or characteristic of the food, such as ingredients, form, processing, or packaging. Names with such a structure explicitly convey information about the food; they are called **explicit** (or expressive) **names**. Examples:

> Sugar-coated puffed wheat
> Chilled spicy tomato soup
> Cut green beans

Most readers understand the component terms and thus can immediately deduce the meaning of these explicit names, even if they have never seen them before. An explicit name encapsulates the food's description.

Other names do not convey such explicit information; they are **nonexplicit names**. Examples:

> Tuff   (a fictitious brand name for sugar-coated puffed wheat)
> Gazpacho
> Surimi
> Hamburger

But how can one have an idea of a food designated by a nonexplicit name? One may know the food by this name; most Americans have an immediate association between the name *hamburger* and a description of that food (*a fried ground beef patty usually between two halves of a roll*). Or one can look up an unknown nonexplicit name in a reference source and find a description of the food. For example, under the name *surimi* one finds the description *imitation shellfish made from minced fish*. A nonexplicit name is the retrieval key for locating a food description in one's memory or in a reference source or database. Most readers know what foods frequently occurring nonexplicit names refer to but must look up nonexplicit names for rare foods.

2.2 Explicit and nonexplicit food codes

Food names and food codes are both used to designate foods. A food name is part of everyday language, such as *tomato soup* or a brand name like *Tuff*. A food code is created specifically to identify and represent a food in a database, such as FP327 for *puffed wheat*. Codes, like names, can be explicit or nonexplicit. **Explicit codes** are constructed from component codes, each of which stands for a food product characteristic; for example, an explicit code for *puffed wheat* looks like this:

$$FP157.SO123.PT035.PR073.$$

The component codes are

FP157	Food Product	Breakfast cereal
SO123	SOurce	Wheat
PT035	ParT	Seed or kernel
PR073	PRocess	Puffing

A person who knows the component codes can readily interpret the combination; thus this code is explicit even though the components codes are readily known only to the few people working

daily with the coding scheme. In contrast, FP327 is a **nonexplicit code** for the same product. The Universal Product Code (UPC) (6) is a system of nonexplicit codes.

2.3 Nonexplicit codes keyed to food descriptions in a database

A nonexplicit name or code has the function of referring to a description; it cannot stand alone. Such a description may be given through a series of statements in a database, as in the following example. In each statement the food *puffed wheat* is identified through the nonexplicit code **FP327**:

FP327	*is a*	FP157 (Breakfast cereal)
FP327	*comes from source*	SO123 (Wheat)
FP327	*comes from part*	PT093 (Seed or kernel)
FP327	*underwent process*	PR076 (Puffing)

These statements about FP327 collectively give a description of the food, and FP327 takes on meaning with reference to them. FP327 refers to this description, this combination of food product characteristics, and can be used as shorthand to refer to this combination in any context. For example, it can be used to refer to the food in a nutrient composition database, where one can then find it in a search for the composition of *puffed products*. FP327 can also be used in the description of a more complex (and more specific) food, such as FP512 *Sugar-coated puffed wheat*. The description consists of the three statements shown in the following diagram:

	is a	FP327 (Puffed wheat)
FP512	*has ingredient*	FP416 (Sugar)
	underwent process	PR213 (Coating)

The first statement links the specific food product FP512 to the more general product FP327. By referring to the statements about FP327, one can infer that FP512 is a *breakfast cereal* (FP157), made from the plant *wheat* (SO123) using the part *kernel* (PT093), and *puffed* (PR076). The specific FP512 **inherits** characteristics from the more general FP327; in artificial intelligence this inclusion of information by reference is called **hierarchical inheritance** ((7), p. 255-264). The other two statements add specific information for FP512: FP512 has the additional ingredient *sugar* (FP416) and underwent the process of *coating* (PR213).

2.4 Levels of specificity of food products and their names or codes

The food products to be designated by names or codes vary in generality/specificity from a **product type**, such as *bread*, to a **generic food product** (which may be standardized), such as *whole wheat bread*, to a **particular brand** of whole wheat bread, to the **individual sample** of whole wheat bread analyzed in a study or the specific slice of whole wheat bread or a sandwich made with whole wheat bread eaten by a respondent in a consumption survey. All of these can

be identified through nonexplicit codes:

FP083 Bread

FP117   Whole wheat bread

FP658   Whole wheat bread, Greentree Farm

S27-574-25   The 25th item on the list of things eaten by respondent no. 574 in food consumption survey no. 27, namely a slice of whole wheat bread, Greentree Farm, toasted

S27-653-12   Again a specific item eaten by a respondent, namely a sandwich made with toasted Greentree Farm whole wheat bread, Fancy Spread salad dressing, and sliced cooked chicken breast

For each of the products so named, there exists a description in the database. With FP083 are associated the characteristics common to all breads. For FP117, there is a statement

FP117 *is a* FP083,

so that FP117 inherits all the characteristics given for FP083. Further statements add specifics for FP117, as in the case of FP512 above. FP658 is treated in an analogous fashion. The specific sandwich eaten has a rather lengthy description:

	is a	FP084 (Sandwich)
	has ingredient	FP659 (Greentree Farm whole wheat bread,toasted)
S27-653-12	*has ingredient*	FP630 (Fancy Spread salad dressing)
	underwent process	PR057 (Spreading FP630 on FP659)
	has ingredient	FP498 (Sliced cooked chicken breast)

If there are many people eating just that kind of sandwich, it is more economical to introduce this specific combination of characteristics as a food product in its own right, identified by a code, e.g., FP923, and associating with it the description

	is a	FP084 (Sandwich)
FP923 ...	...	
	has ingredient	FP498 (Sliced cooked chicken breast)

The record for each item consumed by a person is then shorter:

S27-653-12   is a   FP923

A search for consumption of *toasted products* would find FP923 and then S27-653-12 as an instance of FP923. To give another example, in a food consumption survey one might define codes for frequently consumed fast food hamburgers with brand names. `

2.5 The choice between explicit and nonexplicit names or codes

Two important criteria in the choice between explicit and nonexplicit names or codes are informativeness and the need for brevity. These two criteria are in a trade-off relationship. Explicit names or codes carry their own explanation and are, therefore, long. Nonexplicit names or codes merely refer to an explanation and are, therefore, short. Compare *chilled spicy tomato soup* with *Gazpacho* or

> Sugar-coated puffed wheat with vitamins and iron added, packed in a paperboard box with plastic-coated liner

with

> FP512.

The trade-off must be resolved for human-readable information on the one hand and computer-stored databases on the other. Product labels, survey instruments, texts, screen displays, and indexes to be used by humans should convey information directly without the need for following a reference to another place. As long as this requirement is met, brief names or codes are preferable, because there is less to print or read. This calls for the following rule: If a product has a nonexplicit name that is familiar to the intended audience, use it; otherwise use an explicit name even though it is longer.

Computers, on the other hand, can follow linkages in a database very quickly, and computerized systems should use brief nonexplicit codes to identify foods; as discussed above, these codes serve as keys for accessing food descriptions for
linking descriptive, analytical, and production and use data on a food.

Many existing food databases evolved from manual systems that are now searchable by computer. In such databases foods are usually identified and described by names; there is no further description of foods on which to base retrieval. In this situation explicit names are preferable: A computer search for *tomato* will find *chilled spicy tomato soup* but not *gazpacho*.

While the main purpose of this section was to clarify food names and codes, their relationship to food description and their role in database structure, it already alluded to the more fundamental problems of structuring food description data which is the subject of the next section.

3 The conceptual structure of food description data

In this section we develop a method for structuring and organizing a body of food description data into a systematic, integrated representation that makes explicit the interrelationships between many elements. Such a representation is the basis for information processing and retrieval by humans or by computers.

3.1 The amount of detail in food description

Before discussing the structure of food descriptions, we must reflect on their content, that is, the amount of detail to be included. There is a wide range here. At one end of the scale is a full description giving the amount of each ingredient and time and temperature data for each processing step. At the other end is a superficial description giving just the predominant ingredient. Correspondingly, the form may vary from a narrative or a precisely structured lengthy record in a database to a single word or phrase such as *wheat* or *vegetable soup*.

The amount of detail included in a particular database should balance benefits against costs. More specifically, it should balance the intended uses of the data and the requirements of data exchange against the cost of data collection, storage, and maintenance (8). Detailed descriptions are costly, and the detail is not always justified.

Foods are complex, and a language for the description of foods must be able to express this complexity to the extent required. It should permit description giving any amount of detail, so that the same language can be used by indexers and searchers no matter how much detail is appropriate in a given database. Even broad-brush descriptions should be in the common language so that they can be easily edited by adding detail. Conversely, there should be an automated mapping from very detailed descriptions used in one database to less detailed descriptions required by another. For example, one might envision a computer program that uses detailed recipe information stored in the proprietary database of a food company and derives from it food ingredient and nutrient labeling information.

In this section we describe such a language. If its potential for highly detailed and informative description is used, it requires concomitant sophistication in retrieval made possible by modern computer technology and developments in database management and artificial intelligence. The examples in the next section illustrate this potential for detail. However, the reader should keep in mind that while the language allows for detail it does not mandate it.

3.2 An interlinked database of food descriptions

3.2.1 Basic structure

Food product description is complex because many ingredients and processes are involved. The processes can be performed in locations as different as the manufacturing plant and the home; their sequence is often important. The ingredients are themselves food products which must be described. For instance, consider *chocolate chip cookies* (selected here for its suitability as an

example and not for its nutritional value). The ingredients are *sweet chocolate chip*, which is a form of *sweet chocolate*, which in turn has the ingredients *cane sugar* (extracted from the *stems* of *sugar cane*), *chocolate liquor* (made by roasting and grinding *cacao nibs*, which in turn are made by heating and cracking *cacao beans*, which in turn are the *seeds* of the *cacao plant*), and, in our example, *cocoa butter* (extracted from *seeds* of the *cacao plant*), *lecithin* (may be derived from *soy beans*), and *vanilla* (*ground-up fermented seeds* of the *vanilla plant*).

Table 2 shows how the description of a complex food like *chocolate chip cookie* can be constructed step by step. This **building block approach** handles the complexity of food description by modeling the structure of food descriptions after the structure of the foods themselves: One first describes the simple food products and then uses these descriptions as building blocks in the description of more complex foods. We used this method already in the Section "Nonexplicit codes keyed to food description in a database". In Table 2 description starts with the simple food FP399 *Cacao bean*, which is used in making FP400 *Cacao nib*, which is used in making FP401 *Chocolate liquor*, which in turn is an ingredient of FP461 *Sweet chocolate*, which, in the form of *chips*, is FP462. The description of FP573 *Chocolate chip cookie* now simply refers to the ingredient FP462 *Sweet chocolate chip*, thus incorporating a good deal of descriptive information by reference. This keeps the description of *chocolate chip cookie* manageable without sacrificing detail.

A database of food descriptions organized in this way does not consist of an independent, self-contained record for each food. Rather, individual food description elements are linked to each other, forming an intricate **semantic network** (a term used in artificial intelligence, see (7), Chapter 8, esp. p. 253-257). This elegant, non-redundant storage has its price in retrieval. The complete information about a given food must be pieced together following many relationships through the entire database.

3.2.2 Hierarchical inheritance

There are various types of links between food product descriptions. The **has ingredient** link refers from a product to an ingredient that is already fully described, thereby incorporating its information by reference.
The **is a** link refers to a more general food product, as in

   FP462 (Sweet chocolate chip) *is a*   FP461 (Sweet chocolate)

This reference conveys most of the information needed for describing sweet chocolate chips; only distinctive detail needs to be added:

   FP462 (Sweet chocolate chip) *has form*   Chip

The principle of including information by reference is called **hierarchical inheritance** in artificial intelligence ((7), p. 255-264). FP462 *Sweet chocolate chip* inherits much information from FP461 *Sweet chocolate*, and FP461 *Sweet chocolate* inherits ingredient information from FP401 *Chocolate liquor*. The retrieval system must be able to trace these connections.

3.2.3 The entity-relationship (E-R) approach

The method of food description proposed here is known in the database field as the **entity-relationship** approach ((9), Chapter 3, p. 21-32 and Chapter 9, p. 137-144). In this approach information is structured by considering a set of entities and the relationships that exist between them. Entities and relationships are used to form statements; for example, the statement

    FP399 Cacao bean comes from source   SO287 Cacao plant

links the entities *Cacao bean* and *Cacao plant* by the relationship *comes from source*.

Entities are grouped into **entity types**; the individual entities are called **entity values**. For example, FP399 *Cacao bean* is an **entity value** belonging to the **entity type** *Food product*, and SO287 *Cacao plant* is an entity value belonging to the entity type *organism*. The examples of Table 2 use the entity types and relationship types shown in Table 3. The entity-relationship approach provides a flexible means for structuring complex data, as the examples in Table 2 demonstrate.

The descriptions discussed so far can be amended to include still more information, for example, the quantity or the purpose of ingredients, as illustrated in the following example:

    FP323   *has ingredient*   [SB095 Ascorbic acid, 100 mg in 100 g,
    (Applesauce)   for preservation]

A systematic list of the entity types and relationship types is called the **conceptual schema** of a database. The conceptual schema determines details of the information that can be captured. Table 4 shows a conceptual schema for food description data. The entity types and relationship types listed reflect the types of descriptive information required in the food field. (For good summaries of types of information see (10), (1), (11) (12).) On a more specific level, a conceptual schema also sets forth permissible values for each entity type, as in the standardization of the names of organisms (see the section on hierarchies below). The conceptual schema given in Table 4 can be extended to cover all types of food data.

3.3 Guidance for indexing and searching: frames

The types of entities and relationships given in the conceptual schema define what data **can** be represented. Further instructions to the indexer specify what data **must** be represented. Two examples are:

- A food product **must** be linked directly to another, more general product through the relationship *is a*.

- A product **must** be linked to other appropriate entities through the pair *comes from source* and *comes from part* or through one of the relationships *is made from* or *has ingredient* (unless these are implied by inheritance from a more general food product).

These instructions about the types of information to be included in a food description define a schema or, in the terminology of artificial intelligence, a **frame** with a number of **slots** to be filled (i.e., questions to be answered) by the indexer. (For a discussion of the frame concept, see (7), p. 265-270.) A sample frame is shown in Table 5; this frame represents the structure of FDA's Factored Food Vocabulary (FFV) (1). Other frames may be defined to structure the input of other data.

The systematic list of factors presented in a frame is also helpful in query formulation. In this case, only the slots corresponding to selection criteria are filled; for example, in a search for *Fruit products preserved by drying* only three slots would be filled:

is a	fruit or fruit product
comes from part	fruit or berry
underwent process	drying, for preservation

The frame structure serves to ensure the completeness of the data; it gives the indexer a framework or checklist for the data to be included and reveals missing data as gaps in a pattern. The frame structure also helps the searcher to include all appropriate aspects in the query formulation.

3.4 Hierarchies of entity values

An entity type, such as *organism*, *anatomical part*, or *process*, can have many values (see Table 3). These values should be arrayed in a hierarchy to make it easier for a user to review and comprehend them, and to support efficient aggregative searching (see Table 6).

More specifically, the hierarchical arrangement serves the following purposes:

- Aggregation in searching and data analysis, for example, retrieving all *poultry* as a group, whether *chicken*, *turkey*, *duck*, *goose*, *pheasant*, etc.

- Broad descriptions when specificity is not possible (as in the description of product types) or not desired.

- Imprecise description on a broad level when specific information is not available.

- Guidance in the selection of terms in indexing and searching. A logical arrangement expressing similarities between concepts helps an indexer or searcher to internalize the hierarchical structure.

All decisions in structuring the hierarchy must be made with these functions, especially aggregative searching, in mind.

The base hierarchy of entity values within an entity type should be based on intrinsic relationships among the entities within the type, such as taxonomic relationships among organisms or relationships among chemical substances based on their chemical structure.

Additional hierarchies can and should be constructed based on extrinsic relationships to other entity types, such as the arrangement of plants by usage (as *fruit*, *vegetable*, *spice*) or the arrangement of chemical substances by usage (as *preservative*, *color*, *nutrient*). In the context of a particular application, such an added hierarchy may be more useful than the base hierarchy; the added hierarchy may become the primary hierarchy.

The hierarchies, especially the added hierarchies, are based on data that form databases in their own right, such as the data on the use and harmful effects of chemical substances stored in the PAFA (Priority-based Assessment of Food Additives) database maintained by FDA's Center for Food Safety and Applied Nutrition.

The creation of hierarchies requires considerable effort, but it pays off in terms of improved indexing and query formulation and ease of aggregative searching.

3 Outlook

The flexible and powerful language for the description of foods in databases presented in this paper can serve as the intellectual foundation for sophisticated and user-friendly systems for precise dietary analysis and nutrient intake studies, making full use of the technology pioneered in (13) and (14) and (15) and incorporating expert system features. It supports retrieval and integration of food and nutrition data from many databases. Furthermore it could be used for a formal representation of food standards and of formulations (recipes) used by food manufacturers. Many benefits could be derived from these applications: Design of nutritious yet economical foods, automatic checking whether a formulation conforms with the applicable standards, and computer-assisted production of ingredient and nutrition labeling.

References

(1)   McCann, A., Pennington, J.A., Holden, J.M., Soergel, D., and Wiley, R.C.: FDA's Factored Food Vocabulary for food product description. J Am Diet Assoc 88:336, 1988.

(2)   Herold, P.: Using the factored food vocabulary in food composition databases. Proceedings of the Twelfth National Nutrient Databank Conference, Houston, TX, April 12-15, 1987. Ithaca, NY: CBORD Group, 1987.

(3)   Butrum, R.R., and Pennington, J.A.: Technology systems used for food composition databases: An historical perspective. In Glaeser, P.S., ed.: Computer Handling and Dissemination of Data (CODATA Proceedings). New York, NY: Elsevier, 1987. 404-407

(4)   Harris, L.E., Haendler, H., Riviere, R., and Rechaussat, L. (eds.): International Feed Databank System. Logan, UT: International Network of Feed Information Centers, 1980.

(5)   Haendler, H.: Methods of identifying data units for retrieval purposes, as applied in an international data bank system for feed analyses. (CODATA Proceedings) New York, NY: Elsevier, 1984.

(6)   Hicks, L.E.: The Universal Product Code. New York, NY: AMACOM, 1975.

(7)   Winston, P.H.: Artificial intelligence. Reading, MA: Addison-Wesley, 1984.

(8)   Pennington, J.A.T.: Food composition database features. Proceedings of the Tenth National Nutrient Data Bank Conference, July 22-24, 1985, San Francisco, CA. Springfield, VA: NTIS, 1985.

(9)   Soergel, D.: Organizing information. Principles of data base and retrieval systems. Orlando, FL: Academic Press, 1985.

(10)   Butrum, R.R.: Considerations in designing a system for description of food and flexible retrieval of relevant data. In Glaeser, P.S., ed.: The role of data in scientific process (CODATA Proceedings). New York, NY: Elsevier, 1985.

(11)   Harris, L.E., Butrum, R.R., and Sorenson, A.W.: International food component data systems. J Am Diet Assoc 84:1210, 1984.

(12)   Rizek, R.L., and Butrum, R.R.:The nutrient data bank: A technical information system to aid professionals. National Agricultural Outlook Conference, Washington, D.C. 1975.

(13)   Dare, D., Yarbrough Al-Bander, S.: A computerized diet analysis system for the research nutritionist. J Am Diet Assoc 87:629, 1987.

(14)   Feskanich, D., Buzzard, I.M., Welch, B.T., Asp, E.H., Dieleman, L.S.,Chong, K.R., and Bartsch, G.E.: Comparison of a computerized and a manual method of food coding for nutrient intake studies. J Am Diet Assoc 88:1263, 1988.

(15)   Feskanich, D., Buzzard, I.M., and Biefel, R.: Computerized 24-hour dietary recall data collection for NHANESIII. Proceedings of the Twelfth National Nutrient Databank Conference, Houston, TX, April 12-15, 1987. Ithaca, NY: CBORD Group, 1987.

Table 1. **Food description glossary**

Food name   A name that designates a food; it is part of everyday language. Examples: *chilled spicy tomato soup*, also called *gazpacho*, or *sugar-coated puffed wheat*, a brand of which is called *Tuff*.

Food code   A code created specifically to identify a food in a database, usually some combination of letters and/or digits, such as FP512.

Explicit food name or code   A food name or code built up from component terms or codes, respectively, such that each component expresses a characteristic or aspect of the food. Examples of explicit food names:

> Spicy chilled tomato soup
> Sugar-coated puffed wheat

Example of an explicit food code:

> FP157.SO123.PT035.PR073.PR213.FP416

where

FP157 codes	Food Product	Breakfast cereal
SO123	SOurce	Wheat
PT035	ParT	Seed or kernel
PR073	PRocess	Puffing
PR213	PRocess	Coating
FP416	Food Product	Cane sugar

If one knows the meaning of the component terms or codes, one can immediately infer the meaning of an explicit food name or code, even if one has never seen it before.

Nonexplicit food name or code   A food name or code that is not constructed from components which express characteristics of the food. Examples of nonexplicit food names:

> Gazpacho, Tuff

Example of a nonexplicit food code:

> FP512

The meaning of a nonexplicit name or code cannot be inferred - it must be known or looked up in a reference source or database.

Table 2. **The building block approach to the description of food products**

First descriptions for the fairly simple food products
Cacao bean, *Cacao nib*, *Chocolate liquor*, *Cocoa butter*, *Cane sugar*, *Lecithin*, and *Vanilla* are constructed, forming the building blocks for constructing a description of *chocolate chip cookie*.

	is a	FP017 (Plant product)
FP399 (Cacao bean)	*comes from source*	SO287(Cacao plant)
	comes from part	PT035 (Seed, kernel)
	is a	FP256 (Cacao or chocolate product)
	is made from	FP399 (Cacao bean)
FP400 (Cacao nib)	*underwent process*	PR011 (Heating)
	underwent process	PR076 (Cracking)
	is a	FP256 (Cacao or chocolate product)
	is made from	FP400 (Cacao nib)
FP401 (Chocolate liquor)	*underwent process*	PR173(Roasting)
	underwent process	PR005 (Grinding)
	is a	FP256 (Cacao or chocolate product)
FP402 (Cocoa butter)	*is made from*	FP399 (Cacao bean)
	is extracted substance	SB003 (Fat)
	is a	FP017 (Plant product)
FP415 (Sugarcane stalk)	*comes from source*	SO189(Sugar cane)
	comes from part	PT017 (Stem, stalk)
	is a	FP103 (Refined food product)
	is made from	FP415 (Sugarcane stalk)
FP416 (Cane sugar)	*is extracted substance*	SB005 (Sucrose)
	underwent process	PR 175 (Refining)

FP443 (Vanilla extract)   *(not shown)*

FP457 (Lecithin)   *(not shown)*

Using these building blocks, descriptions for the increasingly complex food products *Sweet chocolate*, *Sweet chocolate chip*, and *Chocolate chip cookie* can be constructed.

FP461 (Sweet chocolate)	*is a*	FP256 (Cacao or chocolate product)
	has ingredient	FP416 (Cane sugar)
	has ingredient	FP401 (Chocolate liquor)
	has ingredient	FP402 (Cocoa butter)
	has ingredient	FP457 (Lecithin)
	has ingredient	FP443 (Vanilla extract)
	underwent process	PR009 (Mixing)
FP462 (Sweet chocolate chip)	*is a*	FP461 (Sweet chocolate)
	has form	PF083 (Chip)
FP573 (Chocolate chip cookie)	*is a*	FP301 (Cookie)
	has ingredient	FP553 (Wheat flour)
	has ingredient	FP416 (Cane sugar)
	has ingredient	FP125 (Vegetable shortening)
	has ingredient	FP203 (Eggs)
	has ingredient	FP462 (Sweet chocolate chip)
	underwent process	PR009 (Mixing)
	underwent process	PR167 (Baking)

Table 3. **Entity types and relationship types used in the description of chocolate chip cookie**

Entity types

Entity type: Food product	Entity type: Anatomical part
Entity values: FP017 Plant product	**Entity values**: PT017 Stem, stalk
FP103 Refined food product	PT035 Seed or kernel
FP125 Vegetable shortening	
FP157 Breakfast cereal	**Entity type**: **Process**
FP203 Eggs	**Entity values**: PR005 Grinding
FP256 Cacao or chocolate product	PR009 Mixing
FP301 Cookie	PR011 Heating
FP399 Cacao bean	PR057 Spreading
FP400 Cacao nib	PR073 Puffing
FP401 Chocolate liquor	PR076 Cracking
FP402 Cocoa butter	PR167 Baking
FP415 Sugarcane stalk	PR173 Roasting
FP416 Cane sugar	PR175 Refining
FP443 Vanilla extract	PR213 Coating
FP457 Lecithin	
FP461 Sweet chocolate	**Entity type**: **Physical form**
FP462 Sweet chocolate chip	
FP553 Wheat flour	**Entity values**: PF083 Chip
FP573 Chocolate chip cookie	
	Entity type: **Substance**
Entity type: **Organism**	**Entity values**: SB003 Fat
Entity values: SO123 Wheat	SB005 Sucrose
SO189 Sugarcane	SB095 Ascorbic acid
SO287 Cacao plant	

Relationship types

Food product 1	*is a*	Food product 2
Food product	*comes from source*	Food source
Food product	*comes from part*	Anatomical part
Food product	*is made from*	Food product
Food product	*is extracted substance*	Substance
Food product	*has ingredient*	Food product
Food product	*underwent process*	Process
Food product	*has form*	Physical form

Table 4. **A conceptual schema for food description data**

Entity types

Food product, recipe, standard	Time(duration)
Organism (species/variety/cultivar of plant or animal) or inorganic food source	Equipment
	Container
Growth stage (maturity)	Place/stage of processing or point in
Environment (e.g., soil type)	distribution chain (e.g., farm, manufacturing
Agricultural treatment	plant, retail store, restaurant, home)
Season	Use/diet
Anatomical part	Consumer group
Cut no. (from permanent plants)	Purpose or effect (e.g.,nutrition, preservation,
Grade, quality	texture, packing)
Substance, material	Meal type
Physical state	Amount
Physical form	Property
Process (incl. storage and handling)	Place (geographic location)
Sequence number of process	Calendar time
Temperature	Money

Food description relationship types (examples)

Food product

is a	Food product 2
is one of	[Food product list]
is analog of	Food product 2
comes from source	[Food source, environment, agricultural treatment, growth stage]
comes from part	[Anatomical part, growth stage, cut, grade]
is made from	Food product
contains	[Substance, amount in total, amount in solids, label claim (yes/no)]
is extracted substance	[Extracted substance, extracting substance, process, temperature, duration, sequ.no.]
had removed substance	[Extracted substance, etc.]
has ingredient	[Food product, rank, total ingredient in
may have ingredient	total product, ingredient solids in product solids [purpose list]]
underwent process	[Process, equipment, temperature, duration, place/stage, sequence no., [purpose list]]
has state	Physical state
has form	Physical form
has property	Property
is for special use	[Use/diet, [country list]]
made for	[Consumer, [country list]]
usually consumed for	[Meal type, [country list]]
contains dish	Food product 2
packed in	Container
has price	Money

Container (similar relationships for equipment)

uses structural strength material	Substance
uses coating material	Substance
has form	Physical form

Thesaurus relationship types (examples)

[Organism, part]	*is used for*	[purpose, priority [country list]]
Substance	*is used for*	[purpose, priority, food product]
Substance	*is harmful for*	[harmful effect, strength, food product]
Substance	*is measured in*	Unit of measurement

Table 5. **A sample frame for the indexing of food products**

Frame slots are defined by a *relationship type* and an entity type (implied by the relationship type or given explicitly) from which the slot values are to be taken. Each slot has a **name**

Food product

A *is a* (product or **Product type**) _____

A1 *comes from organism* (**Food source)** _____

A2 comes from part **Part of plant or animal** _____

   OR, replacing A1 and A2

   is made from OR *has ingredient* _____

C *has form* **Physical state, shape, or form** _____

D *underwent process*

D1 underwent **Degree of preparation** _____
   (How much heat was applied)

D2 *underwent* **Cooking method** _____
   (A subdivision of processes)

D3 *underwent* **Treatment applied** _____
   (All other processes)

D4 had applied **Method of preservation** _____
   (*has ingredient* or *underwent process*)

E1 *has ingredient* as **Packing medium** _____

E2 *is packed in* **Container, wrapping** _____
   which has main material and form

E3 container has **Food contact surface** _____
   (container *uses coating material*)

F *made for* **Consumer group** _____

Table 6. **Sample hierarchies of entity values**

Organism by taxonomy
Plant

Cruciferae
    Brassica
        B. oleracea
            var. gemmifera - Brussels sprouts
            var. capitata - Cabbage
        B. nigra - Black mustard
        B. hirta - White mustard

Leguminosae
    Faboidae (subfamily)
        Vicia
            V. faba - Broad bean
        Phaseolus
            P. vulgaris - Common bean
        Glycine
            Glycine max - Soybean
    Caesalpinioidae (subfamily)
        Ceratonia
            C. siliqua - Carob

Byttneriaceae
    Theobroma
        T. cacao - Cacao plant
Sterculiaceae (closely related)
    Cola
        C. acuminata - Cola tree

Theaceae
    Thea
        T. sinensis - Tea

Rubiaceae
    Coffea
        C. arabica - Common coffee

Organism by use
Plant

Plant producing extract or concentrate
    Oil-producing plant
        Soybean (Glycine max)
        Cacao (Theobroma cacao)
    Protein-producing plant
        Soybean (Glycine max)

Spice- or flavor prod. plant
    Mustard
        Black mustard (Brassica nigra)
        White mustard (Brassica hirta)
    Cacao-flavor producing plant
        Cacao (Theobroma cacao)
        Carob (Ceratonia siliqua)
    Coffee-flavor producing plant
        Common coffee (Coffea arabica)

Plant used as stimulant
    (most are also flavor-producing)
    Cacao (Theobroma cacao)
    Coffee (Coffea arabica)
    Cola (Cola acuminata)
    Tea (Thea sinensis)

Vegetable-producing plant
    Head vegetable
        Brussels sprouts (Brassica
            oleracea, var. gemmifera)
        Cabbage (Brassica oleracea, var. capitata)
    Pod or seed vegetable
        Bean
            Broad bean (Vicia faba)
            Common bean (Phaseolus vulgaris)
            Soybean (Glycine max)

Anatomical part
Part of plant

Root, stem, leaf, or flower
    Root, tuber, bulb
    Above-ground part
        (excluding fruit or seed)
        Stem, stalk, shoot
        Leaf
        Head
        Floret or flower
            Calyx
            Stigma

Fruit or seed
    Fruit or berry
    Pod or seed
        Pod
        Seed or kernel
            Skin (bran) present
                Skin present, germ present
                Skin present, germ removed
            Skin (bran) removed
                Skin removed, germ present
                Skin removed, germ removed (endosperm)