Framework of Nutrient Information by Using Facet Classification

Wai Mar Tint, Mi Mi Nge
University of Computer Studies, Yangon
wmcko30786@gmail.com, dawmiminge@gmail.com

Abstract

Effective information architecture (IA) is mapped to the structure of actual human systems. The IA of an information system must be hospitable to change as well. The representation of actual human knowledge is captured by taxonomic structures. A faceted classification system (FCS) which is essentially a multidimensional taxonomy. FCS can mimic more closely the structure of actual knowledge. Faceted classification is suitable for organizing digital information resources. Since a nutrient is a substance obtained from food and used in the body to promote growth, maintenance and repair of body tissues, everyone needs nutrient information. In this paper, nutrient information architecture is implemented by using facet classification. This paper provides search and browse for nutrient information retrieval and also supports query formation and expansion by browsing. In this paper, the system for nutrient information architecture is implemented based on XFML.
Keyword: Facet, FCS, XFML

1. Introduction

Real human knowledge is, by nature, inconsistent, ill-defined, and unstructured. It is indescribable, impossible to represent faithfully without some loss of meaning. As a result, any method of knowledge representation is flawed. The products of human knowledge lie in documents, and all attempts to describe or organize those documents are flawed as well [1].

Digital libraries or web sites often rely on subject hierarchies (such as Yahoo) to arrange documents for browsing. However, many documents are "about" more than one thing. This often leads to either duplication of documents or the limiting of access to one list of subjects. The former leads to wasted time and resources. The latter removes possibilities for meeting users' needs [9].

Users employ several information seeking and retrieval strategies to find what they need, including browsing, known site, empirical, analytical, and similarity. Unfortunately, many information architectures only allow for a few of these [4].

Real knowledge is multi-layered, where meaning is synthesized from several different perspectives. It is infinitely multi-dimensional. The intersection of more perspectives on knowledge leads to higher relevance, but few existing information systems include this functionality [2].

Real knowledge is chaotic, disorganized, and constantly changing. Document change as well, and their meaning changes as society changes. As the shape of the knowledge changes, so must the structure of the systems used to access that knowledge [5].

Section 2 describes the related work with the classification system. Section 3 briefly introduces the faceted classification system. In section 4, XFML language is used to put faceted classifications into a standard and human readable form. In section 5, the proposed system design of nutrient information architecture is implemented. In section 6, the implementation about the framework is described. Finally, conclusions are presented in section 7.

2. Related work

S.R. Ranganathan pioneered a method of classifying documents using a multi-dimensional notation called Colon Classification. By organizing documents according to several subjects simultaneously, the shape of knowledge can more faithfully be rendered in the shape of the collection.

The basic of this approach is the adoption of a particular system for classifying the collected items: facet classification (also called analytico-synthetic classification or multidimensional classification) [8]

In this way, a user is able to choose whether to look for information by direct search or by browsing a directory. When, for the purpose of the classification, it is possible to organize the entities by three or more mutually exclusive and jointly exhaustive categories, then facets are probably the appropriate classification.

In traditional classifications systems, every element is classified under one and only one class. It has a unique position inside a hierarchical and very deep schema, and it can be found through a unique path: father

category > child category. Dewey decimal classification and the Library of Congress Classification are examples of hierarchical-enumerative systems. [12] Schemas they based on are very rigid and conservative; because they are structurally closed, and required a centralized management organization.

2.1. **Hierarchical-Enumerative**

A taxonomic top-down scheme, in which knowledge is divided into progressively narrower and more specific categories (a hierarchy). Enumerative classification assigns names to every subject and enumerates them, typically in a systematic order. When this approach is applied to libraries, where physical objects -- books -- are classified, each object is located in a single place, often with one only path to get there. Hierarchical enumerative structure is shown in figure 1.

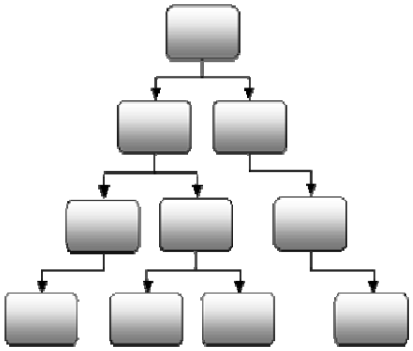

Figure 1: Hierarchical-Enumerative Structure

In contrast, faceted systems avoid the need of listing all the possible classes of the taxonomy, and allow instead creating the classes by combining elementary elements (facets and foci). Besides, they allow associating an item to several categories or parameters at one time, each category representing an aspect or face of the item itself. [7]

2.2. **Analytical-Synthetic**

A faceted bottom-up scheme, that breaks down a subject into individual concepts (*analytical*) and provides rules to use these concepts in constructing headings for composite complex subjects (*synthetic*). New elements can be developed as new concepts emerge, often without superseding the previous categorization activity. The principal characteristics of a faceted system can be summoned up as follows:

i. **Multidimensionality**: in spite of the traditional systems, in the faceted systems each object is classified according to different attributes (called facets)
ii. **Persistence**: these attributes/facets are essential and persistent properties of the object; in this way the impact (on the classification system) of eventual changes (of nomenclature, of workflow and so on) is reduced or is insignificant;
iii. **Gradualism**: it is always possible to add a new descriptive facet of a new aspect of the object;
iv. **Flexibility**: there is a variety of parallel access keys (facets); every object can be found using one only attribute (facet) at time or different attributes combined.

Faceted classification results really flexible and extensible, allowing everyone to create own classes. Such systems are especially useful in managing a large amount of items. Figure 2 shows the multi-dimensional structure of a faceted scheme.

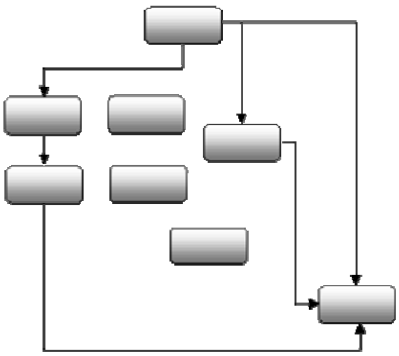

Figure 2: Multi-Dimensional Structure

Example, a faceted list of fruits could be as follows:

Name: Apple, Orange, Banana, etc.

Color: White, Red, Yellow, etc.

.

Vitamin: Vitamin A, Vitamin B1, etc.

Mineral: Calcium, Potassium, Iodine, etc.

Looking for a fruit, a user can choose among several query parameters; alternatively combine these parameters in order to restrict the range of the results. For instance, a user can select only fruits having characteristics like: Name: Orange; Color: Yellow; Vitamin: Vitamin C; results will be in a limited number.

3. Faceted Classification

Faceted classification systems assign to subjects clearly defined, mutually exclusive, and collectively exhaustive aspects, properties or characteristics of a class or specific subject. Faceted classification provides more than one path to locate a subject. By using different aspects/ facets of a subject, users can narrow down the search and locate the subject more easily. Although each of hierarchical and faceted systems has its own strengths and weaknesses, traditional hierarchical structure is increasingly seen to be ineffective in organizing resources on the web.

Faceted classification is flexible and hospitable in accommodating new categories. The faceted approach is more suitable in representing complex ideas/objects. It is particularly useful for organizing complex items and materials in a multi-disciplinary environment [3]. A faceted approach, flexible and hospitable to multiple perspectives, should be more suitable for classifying cross-cultural and multilingual resources. [6]

Faceted classification is also called analytico-synthetic. It is named after the two main processes involved in the composition. The two processes are:

Analysis: Breaking down each subject into its basic concepts.

Synthesis: Combining the relevant units and concepts to describe the subject matter of the information package in hand.

4. XFML

XFML is a markup language written in XML, and hence looks similar to HTML. It is used to put faceted classifications into a standard machine- and human-readable form that is easy to store, transmit and manipulate. There are two main elements in XFML: facet and topic. The facet element defines the top-level facets. It has only one attribute, id, which will be the name used internally to identify the facet. It can be an abbreviation or code number.

Sample of XFML:

<facet id=”name”> Name </facet>
<facet id=”color”> Color </facet>
<topic id=”red” facet id=”color”> <name> Red </name></topic>
<topic id=”white” facet id=”color”> <name> White </name></topic>

Once the whole classification has been rendered in XFML, there remains the problem of how to use it. There seems to be little open source software for XFML available on the Internet and there are no full-featured XFML libraries under continuing maintenance for any major programming language. However, because XFML is written in XML, any of the many XML libraries can be used to handle it. There is one commercial product that handles XFML, facet map.

5. Framework of Nutrient Information

This paper describes procedures for making the faceted classification system and adds to the start and finish to make it complete from beginning to end.

1. **Domain collection.** Collect a representative sample of the entities. In a large domain, get enough to cover all foreseen possibilities. In a small domain, use the entire domain. In this paper, fruits type - 5o, vegetables type - 42 and nuts and seeds types -15 are collected for this domain.
2. **Entity listing.** List the entities, breaking down the descriptions into parts and rearranging words. Separate sentences and phrases into their basic concepts and isolates these concepts. For example, name, color, vitamin, mineral are specifically listed in this step.
3. **Facet creation.** Narrow them down into a set of mutually exclusive and jointly exhaustive facets. Examples of facets are: Name (Apple, Avocado, Blueberry); Color (Blue, Red, Yellow); Vitamins (Vitamin A, Vitamin B1, Vitamin B2); Minerals (Calcium, Potassium, Sodium).
4. **Facet arrangement.** First, do a test ordering of the terms (now to be known as “foci”, or individually, “a focus”) under the facets. Second, after the rough draft of the classification has been shown to work, do the final facet arrangement, using the Principles for the Citation order of Facets and Foci. That arranged the foci under the previous categories in alphabetical order.
5. **Citation order.** Choose a primary facet that will determine the main attribute and a citation order for the other facets. This step is not required and applied only in those situations where a physical organization is desired.
6. **Classification.** The classification system is finished. Use it to classify

everything in the domain. Analyze the entities using the facets, and apply the proper foci to describe each entity.

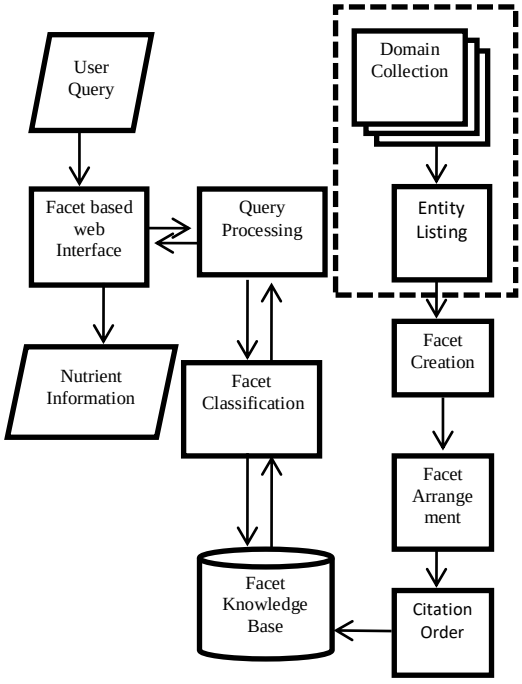

Figure 3: System Design of Nutrient Information Architecture using Facet Classification

The overall system includes six modules. The first module is domain collection which collects the numerous data concerned with the nutrient domain including food sources such as fruits, vegetables, nuts and seeds. Second module is entity listing. In nutrient domain, the fruit names are already broken down into constituent concepts such as name, color, vitamins, minerals, etc. This makes the facet creation, easier. In facet creation, high-level categories appear across all the entities. Some stand out: Name (Apple, Avocado, Blueberry); Color (Blue, Red, Yellow); Vitamins (Vitamin A, Vitamin B1, Vitamin B2); Minerals (Calcium, Potassium, Sodium), as does the principle of Mutual Exclusivity.

Next module is facet arrangement that arranged the foci under the previous categories in alphabetical order. The default citation order is name, color, vitamins, minerals, benefit symptoms and deficiency symptoms, but the users can reorder the facets to their liking. And then, classification is doned by analyzing the entities using the facets.

Nutrient information is stored in XML format in the faceted knowledge base. Query processing makes easier for users to find what they are looking for. If the user wants to know the contents of specific nutrients, he/she can easily get information by using facet based web interface.

6. Implementation

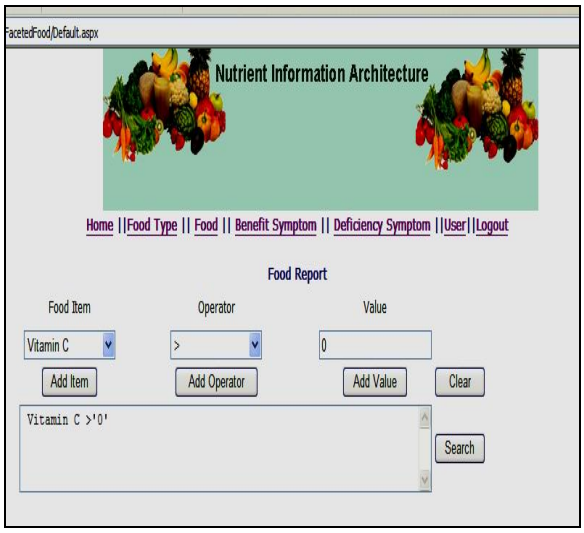

Figure 4: User interface to accept criteria

In this paper, the datasets are collected as follows: fruit types- 42, Vegetables types- 48, Nuts and Seeds types- 16, Vitamins types- 12, Minerals types- 11, Benefit Symptoms- 10and Deficiency Symptoms- 9. The facetID, facet name, quantity and resource of URLs are described as shown in Table1.Figure 4 shows the web interface of entering criteria to retrieve information according to user requirement, user need to select the attributes of the food.

This paper is an implementation of framework for Nutrient Information Architecture using Faceted Classification Approach. A faceted scheme of classification is implemented based on (XFML) exchangeable faceted metadata language.

Traditional classificatory structure defines the relationship of contents in a classification scheme. In a hierarchy, the whole field is divided into classes and each class can be divided into subclasses. All the divisions are based upon the generic relationship of knowledge. A tree divides the information into classes and subclasses in the same way as a hierarchy does. The main difference is that all the divisions in a tree can be based upon any

Table1: Collection of Nutrients, Benefit and Deficiency Symptoms

Type	Facet ID	Facet Name	Amount	URL
1	Fruits	Apple, Avocado, Banana, Blackberries, Blackcurrants, Blueberries, Boysenberries, Breadfruit, Cantaloupe, Cherimoya, Cherries, Chinese pear, Cranberries, Dates, Figs, Gooseberries, Grapefruit, Grapes, Guava, Kiwi, Lime, Loganberries, Lychee, Mango, Mulberries, Nectarine, Olives, Orange, Papaya, Passion fruit, Peach, Pear, Persimmon, Pineapple, Plum, Pomegranate, Prickly pear, Raspberries, Starfruit, Strawberry, Tomato, Watermelon	42	http:// www.healthalternatives2000.com/fruit-nutrition-chart.html
2	Vegetables	Alfalfa, Amaranth leaves, Artichoke, Asparagus, Bamboo shoots, Beetroot, Bokchoy, Broccoli, Brussel sprouts, Butternut squash, Cabbage, Carrots, Cauliflower, Celeriac, Celery, Chinese broccoli, Chinese cabbage, Corn, Cucumber, Daikon raddish, Eggplant, Fennel, French beans, Green pepper, Jicama, Kale, Leek, Lima beans, Mushroom, Okra, Onions, Peas, Potatoes, Pumpkin, Parsnip, Radish, Rapini, Spinach, Spirulina, Spagetti squash, Squash summer, Squash winter, Sweet potatoes, Swiss chard, Taro, Turnip, Yellow squash, Zucchini	48	http:// www.healthalternatives2000.com/fruit-nutrition-chart.html
3	Nuts and Seeds	Almonds, Brazil Nuts, Cashews, Chestnuts, Coconut, Flax seed, Hazelnuts, Macadamias, Peanuts, Pecans, Pine nuts/Pignolias, Pistachios, Pumpkin Seeds, Sesame seeds, Sunflower seeds, Walnuts	16	http:// www.healthalternatives2000.com/fruit-nutrition-chart.html
4	Vitamins	Vitamin A, Vitamin B1, Vitamin B2, Vitamin B3, Vitamin B5, Vitamin B6, Vitamin B9, Vitamin B12, Vitamin C, Vitamin D, Vitamin E, Vitamin K	12	http:// www.healthalternatives2000.com
5	Minerals	Calcium, Copper, Iodine, Iron, Magnesium, Manganese, Phosphorous, Potassium, Selenium, Sodium, Zinc	11	http:// www.healthalternatives2000.com
6	Benefit Symptoms	Bone Growth, Cell Reproduction, Formation of hormones, Function of digestive system, Function of nervous system, Healthy skin, Red blood cells production, Stimulates immunity, Tooth development,	10	http://www.nutristrategy.com/nutritioninfo2.htm
7	Deficiency Symptoms	Confusion, Convulsion, Dizziness, Dry skin, Fatigue and weakness, Nausea, Night blindness, Poor bone growth, Weak tooth enamel	9	http://www.nutristrategy.com/nutritioninfo2.htm

type of systematic relationships instead of only the generic relationship in a hierarchy. A paradigm describes the entities from two perspectives. The faceted classification views the field from multiple perspectives. In such a structure, the whole field can be regarded as a diamond and it can be divided into classes from each angle.

Advantages of faceted classification system are: it does not require complete knowledge of the field; it is hospitable and flexible; it is expressiveness; it is able to accommodate a variety of theoretical structures and models and multiple perspectives. It also has disadvantages: it is difficult to establish appropriate facets; lack of relationships among facets and difficulty of visualization.

XFML is a simple XML format for exchanging metadata in the form of faceted hierarchies, sometimes called taxonomies. XFML provides a simple format to share classification and indexing data. It also provides two ways to build connections between topics. It's based on the principles of faceted classification, addressing many of the scaling issues with simple hierarchies. Sample of XFML tags are described below:

<NewDataSet>

<Table1>
<FoodID>1</FoodID>
<FoodName>Apple</FoodName>
<Color>Red</Color> <TypeID>1</TypeID>
<V_A>2</V_A> <V_B1>0.02</V_B1>
<V_B2>0.01</V_B2>
<V_B3 /> <V_B5 /> <V_B6>0.05</V_B6>
<V_B9>4</V_B9>
<Calorie>44</Calorie>
<Carbohydrate>10.5</Carbohydrate>
<Sodium>2</Sodium>
<Calcium>4</Calcium>
<Potassium>121</Potassium>
<Phosphorus>11</Phosphorus>
<Iron>0.2</Iron> <Magnese />
<Copper>0.02</Copper>
<Zinc>0.04</Zinc> <Protein>0.4</Protein>
<Fiber>2.3</Fiber> <Water>84</Water>
<Fat>84</Fat>
</Table1>
<NewDataSet>

Example of Food XFML

Some other information such as Benefit Symptoms and Deficiency Symptoms are also available in this system. Benefit Symptoms are: Bone Growth, Cell Reproduction, Formation of hormones, Function of digestive system, Function of nervous system, Healthy skin, Red blood cells production, Stimulates immunity, Tooth development and Vision. Deficiency Symptoms are: Confusion, Convulsion, Dizziness, Dry skin, Fatigue and weakness, Nausea, Night blindness, Poor bone growth and Weak tooth enamel. There are two menus at the left hand side of the home page, "Benefit Symptoms" and "Deficiency Symptoms".

The selection of criteria for Benefit and Deficiency Symptoms are different from the criteria entry of nutrients. Output results are described as: Food name: Orange, Peach, Tomato, Watermelon, Asparagus, Broccoli, Carrot, Green pepper, Kale, Peas, Avocado, Spinach, Squash-summer, Sweet potatoes, Almonds, Chestnuts, Pecans and Pistachios. The result nutrients related to the selected symptoms are shown in the same of format of nutrients information.

7. Conclusion

The Internet lacks a systematic structure. Classifying and cataloguing Internet documents is a step forward in organizing it. A faceted classification scheme is more suited to the dynamic nature of the Internet. Due to its flexibility a faceted scheme can keep pace with the ever increasing information on the web and also the rapid emergence of new topics. In this paper, the system for nutrient information architecture is implemented based on XFML. Any interested person can easily get information about food sources such as fruits, vegetables, nuts and seeds by using the faceted classification system.

References:

[1] B.H.Kwasnick,. "The role of classification in knowledge representation and discovery". *Library Trends* 48 (1): 1999

[2] B.Y.Ricardo, and R.N.Berthier, *Modern Information Retrieval*. New York: ACM Press 1999

[3] Broughton, "Faceted classification as a basis for knowledge organization in a digital environment". *The New Review of Hypermedia and Multimedia* 2001

[4] com/cgi-bin/giftfinder .pl November 2003.

[5] J. Douglas. "Facet analysis". *In Encyclopedia of Library and Information Science*, 2nd ed., ed. Miriam A. Drake (New York: Marcel Dekker), 2003

[6] L.VernonOnline.http://www.lillianvernon.

[7] P. A. Heidi, "Use of faceted classification".http://www.webdesign

[8] P. S. Chen, "The entity-relationship model toward a unified view of data". *ACM Transactions on Database Systems* (1) (March 1976)

[9] http://www.practices.com/navigation/facets.htm (November 2003)

[10] R. Louis, and P. Morville. *Information architecture for the World Wide Web*. 2nd ed. Sebastopol, CA: O'Reilly 2002.

[11] S. Louise, "A simplified model for facet analysis: Ranganathan 101". *Canadian Journal of Information and Library Science* 23 (1/2) (April-July) 1998.

[12] V.D.Peter, "XFML Core eXchangeable Faceted Metadata Language". http://www.xfml.org/spec/1.0.html (accessed 15 October 2003).

[13] Vickery, *Faceted classification: a guide to construction and use of special schemes*. London: Aslib. , B.C. 1960

[14] W. Denton. "Putting facets on the web: an annotated bibliography". http://www.miskatonic.org/library/facet-biblio.html. 2003