Currently, faceted classification goes into the realm of the Internet “information retrieval” and it is more accepted as a tool outside the realm of the library and information science. Unfortunately, IT people have adopted the faceted classification without a methodologically full understanding of faceted classification. There is not much theoretical discussion about the possibility on the Internet environments based on its library original. In this paper, I will discuss the potential usage of faceted classification for “information retrieval” from a fundamental problem of relation of classification and searching/retrieval in an organic “information retrieval” system, which has much influence with changing from the library environments to the Internet environments. This change gives chances and challenges to faceted classification. With this change of environments, I want to argue that faceted classification keeps the merits of classification in information retrieval comparing with search engines and overcomes the problems with a traditional enumerative classification structure by providing for the flexible combination of terms to represent a specific complex/compound subject. Especially I want to argue that expressiveness of notation in faceted classification, which has been thought of as a defect until now, can find new usage as a potential switching/connection to multilingual information and a significant aiding in browse searching/retrieval and keyword searching/retrieval.

 KEYWORDS: faceted classification, enumerative classification, information retrieval, shelf organization, expressive notation

1. Introduction

Currently, faceted classification goes into the realm of the Internet information retrieval and it is more accepted as a tool in Knowledge Management and Information Architecture [La Barre, 2004]. Unfortunately, there is hardly any theoretical discussion about the possibility of faceted classification on the Internet environment based on the original theory developed by Ranganathan and the Classification Research Group (CRG) in the realm of IT people, for example, on Faceted Classification Discussion mailing list.1 It seems that it has been partly due to a failure to dig down the roots of the methods. A basic study of methodology has been the factor taken for granted. It is this task that I have here attempted to initiate. This paper will retain its original character as a study of methodology; classification and searching/retrieval used in the paper are methodology of organization and searching/retrieval technology in the organic “information retrieval” system.

I will clarify the basic term of “information retrieval” as an organic system. I will focus especially on the relation of classification and searching/retrieval in it. Especially, as the Internet expands, the task of “information retrieval” has become increasingly problematic and complex. A great deal of disagreement exists as to the individual definition of the terms and there may also be subtle differences from one example to another. For the purpose of this paper, “information retrieval” has its base on definitions of “information retrieval” by F. Wilfrid Lancaster, which is also quoted by Van Rijsbergen [Rijsbergen, 1979]; Baeza-Yates and Riberio-Neto. [Baeza-Yates, Riberio-Neto, 1999]. It is defined as a system, broadly means the whole process of five steps; collection, organization, storage, searching/retrieval and usage of information. The definition of the term “classification” in this paper is based on this concept.8 It is used in this paper as a tool of information organization on the second step of the whole “information retrieval” system.9 “Searching/retrieval” is narrowly defined as retrieval technology, based mainly on the search strategy it employs. In the realm of the Internet, “searching/retrieval” narrowly means full-text search provided by search engine technology.10 All the argument will be based on this change, which gives chances to faceted classification. I will start with a thorough research of the faceted classification in terms of the library environments and the Internet environments, considering different meanings under the two information carriers. I will consider how such meanings influences faceted classification.

2. Problems Encountered

Development of the Internet has lengthened the capacity of search engines to deal with searching/retrieval of information. The advent of search engine technology has actually made a big step forward in the quality of searching/retrieval. It soon becomes the main-stream on the Internet and traditional information organization method, such as
classification, is dumped. People rush to hold close to the straightforwardness of full-text search provided by search engine technology.

However, the incompetence of machines would never take the place of the analytical brainwork. It is only a means to aid and assist, but in no way can they depose the supremacy of the human intellect. We must derive all the benefits capable of giving to classification but this does not mean substitute it for retrieval technology. Many researchers already repeatedly expressed such idea long ago, such as B. C. Vickery in 1960s [Vickery, 1960] and A. P. Srivastava in 1970s [Srivastava, 1977].

In despite of all these, regretfully, we still tend to put all bets on technology every time new technology appears. In answering the question of "why it is so hard to find information on the Web, and why search engines aren’t more helpful!", Rosenfeld replied that "one reason for this confusion is that it’s really hard to express our information needs in words, much less translate those words into a query language understood by a dumb piece of software (that is, search engines). Another reason is that it’s really hard to index the ideas and concepts that are stored in text in a way that this dumb software can understand (and therefore find). So when we do a search, we’re asking something much dumber than we are to do something we find hard to do ourselves" [Morville, 2002].

The fundamental problem here is that the idea of almighty search engines is focused only on the fourth step of "information retrieval" as I have defined above. It seems that search engines are presented as alternative techniques which do away with any need for classification. Classification methods are expelled as outsiders, as if search engines are brand new to serve all the duties for Internet "information retrieval". People did not realize Internet "information retrieval" as an organic system, which is far beyond the reach of pure search engine technology. They thought of classification and retrieval as conflict factors and questioned the value of classification, still far from giving proper evaluation of different types of classification as enumerative classification and faceted classification.

The increasing use of the Internet for the storage and retrieval of vast amounts of information has, however, changed this view. Rather than endless reinvention of seemingly novel solutions, there is the need to find the key for searching/retrieval in a more expansive realm outside of search engines realm themselves. So there are rehabilitated interests in classification. People find that to be faced with information collection and not to be able to find the information that you know exists somewhere is a problem as old as the existence of physical document collections in the libraries. Only in the library environments, it is not as caustic and imminent as the Internet environments. A lot of classification efforts appear to be put into effect, mixtures of various ways/methods in a variety of approaches. But the fact is that the problem of Internet "information retrieval" still can not be mitigated.

The reason is that unfortunately library techniques cannot be said to have advanced fast enough to keep pace with the demands made on them and we have seen during recent years how research workers, particularly in science, have become agitated to the point of calling a number of international conferences to discuss these matters soon after World War II until the appearance of the Internet, especially those efforts by CRG, which is almost the sole research groups doing faceted classification after Ranganathan. Yet little has emerged because classification and retrieval have not really been well realized in an organic system in those researches and IT people and library people have never really made coordination in cooperative researches. IT people cannot be particularly familiar with library techniques, and the same is on the contrary to library people and neither can spend much time in discussion on a very commonplace level on the other side. The situation is slowly improving under the pressure of necessity, but IT people and library people seem to remain out of touch over classification and searching/retrieval in cooperation, simply because they do not think classification and searching/retrieval in the whole "information retrieval" system.

Therefore, in this paper, I want to argue that not all classification schemes are suitable for Internet "information retrieval" from the aspect of classification. The circumstances are changed. A proper classification scheme in the libraries does not exactly fit the Internet environments. In order to avoid the argument in an isolated way, my stance on faceted classification for Internet "information retrieval" is in the organic "information retrieval" which has gradually changed from the library environments into the Internet environments.

3. Faceted Classification in Two Environments

As information carriers changes from the library environments into the Internet environments, some intrinsic problems of faceted classification in the library environments should be paid attention to, when the possibility of the Internet application is taken into consideration.

First of all, "information retrieval" is not a new field. "The retrieval process begins when a lack of information shows itself in a human mind and the decision is taken to find out if this information has been discovered and published" [Foskett, D. J., 1974]. At bottom, it is still nothing more, as a simple word labeling of documents or parts of documents so that information can be more quickly found when it is wanted. While with no doubt such a service has long existed in the libraries. In the libraries, the role of classification in relation to searching/retrieval is expressed as that "the purpose of a classification scheme is to arrange information, in documents on shelves or on cards in indexes, in a sequence that will be helpful to the user" [Foskett, D. J., 1974]. Shelf organization’ plays a most important function, in which the idea of “one book, one place” is very central to it.

In one way, knowledge itself is multi-dimensional. One book can deal with one subject from several different points
of view. Theoretically all these aspects should be identified if needed. But the materialized forms of knowledge are restricted to a linear, one-dimensional sequence, simply for the arrangement of books has to proceed along a line only and in one direction only. Therefore, although enumerative classification has the principle to list and enumerate all the lists in order to fulfill its function, it has an inner problem that not all the details can be enumerated [Ranganathan, 1963]. For the purpose of shelf organization, it should always be realized that highly specialized and minute analysis, arranging and recording of information, which will be complex and bring trouble to shelf organization. If classification is too fine, spaces must be left on the shelves at very frequent intervals, perhaps after every book, to accommodate prospective accessions. Enumerative classification can be suitable and economical for a closed information system and they account for early years of prosperity of the libraries since the middle of 19th centuries.

While it results in an inconsistency with the problem that must distort all knowledge in its infinite multidimensionality into a linear arrangement suitable for creating a browse-able list or locations on shelves. This makes a sacrifice of multidimensional characteristic of information. If we want the books in one facet to come together, those in another facet lose its unity. Library classification is equal to a representation of a multi-dimensional continuum in a single one-way dimension [Ranganathan, 1965]. And the progress of knowledge itself fastens this inconsistency and makes it more and more violent. As more and more are discovered and published, the result is a narrowing of range and more specialization, which in turn brings a tendency to concentrate on details in a new field, and discovery is made in small units more suitable for publishing as essays than as books. The searching unit began to change from “book” to “a piece of article in the periodical”.

Gradually, the function of searching/retrieval (directly for information) began to be isolated from shelf organization (the information carriers, books) when searching for a piece of information. “Information retrieval” has developed as an important independent part. Obviously, with the progress of computer technology since 1950s, it was sometimes claimed that classification could no longer attempt to do anything but arrange books in a more or less convenient order. It meant that there was no classification in the libraries that could offer enough advantages to make the undertaking worthwhile for “information retrieval”. I have made an attempt to address myself to faceted classification itself, in the hope of clarifying the role of it in Internet “information retrieval”.

During the time of invention of faceted classification by Ranganathan, from the aspect of the information carriers, the foundation of classification began to be under pressure of changeover from arrangement of relative macroscopic masses of information embodied in books to relative microscopic units of information embodied in articles. Ranganathan envisioned a way, facet analysis in which we can dynamically organize information storage. In faceted classification, only available choices are visible, which are highly interactive in express compound/complex subjects. The idea was that you could parallel group entities based on various characteristics.

Some interesting story will be very helpful for an easy understanding of his theory. It said the idea about facet stems from Ranganathan’s contact with a Meccano toy set at Selfridges, the London department store. There he saw an entirely new toy with combination of nuts bolts and strips. He found that by combining the pieces in different ways, many different objects could be constructed. The experience made Ranganathan realize that his classification scheme should as well consist of elements that could be freely combined to meet the needs of each specific subject. In classification, by combining the classes in different unit schedules in assigned permutations and combinations, all possible topics can be made. This idea is called “Meccano principle” [Srivastava; 1977]. In his classification scheme, in order to connect the various facet notation symbols, he introduced the colon (:) and other punctuation marks in place of the bolts and nuts of the Meccano set, so the classification scheme is named as Colon Classification (cc). Here, I will just give a general description of it. At a first glance, the CC may look like a simplified version of other classification schemes with enumerative structure; it begins with a list of main classes, each followed by subdivisions. But the point is the introduction of facet. A facet is an atomic unit of collection standing for a viewpoint, a perspective, or a dimension from which to describe a factor. Put in everyday language, a concept is composed with many characteristics. A facet gathers one characteristic of all concepts that have the same characteristic. The concept of gender, for example, partitions a population of persons into two: males and females. Quoting from Vickery, the essence of facet theory is “a method of controlling the kind and level of term that is admitted into the system vocabulary” and is about “the sorting of terms in a given field of knowledge into homogeneous, mutually exclusive facets, each derived from the parent universe by a single characteristic of division.” [Vickery; 1968] [Vickery; 1960].

Compared to an equivalent enumerative classification structure, it is much shorter in length; the list can be so short because it contains only the bits and pieces (facets), not the compound terms. We can find that in fact, enumerative classification is just enumeration from language facet. Faceted classification differs in that the distinguished facets are not fixed into firm, enumerative lists, but are free to combine with one another [A. C. Foskett; 1969]. So by combining facets in complex subjects, all the logical relations among them can be expressed, thus it will better reflect the complex multi-dimensional information of different user needs.

From the aspect of contents of information, what distinguishes the universe of information is a dynamical continuum. It is forever growing, new branches may stem from any of former at any time, and they can not therefore be enumerated here and now and their filiations can be determined only after they appear. Nothing can there be said to be peculiar to one context, and in classification this means that nothing can be considered as peculiar to a single class, and only to be found in that class. Originally, many-faceted subjects are characteristic of the literature; the users’ interest may be
4. Make the Most of Faceted Classification

With the liberation of faceted classification from the limitation of the library environments, I want to give a special argument on the important characteristics as analytico-synthetic can work much better than in the library environments. Proper classification for the Internet should meet the following criterions: It must support finding components that are similar, not just matches the same without contexts; It must be very precise and have high descriptive power; It must be easy to maintain, that is, add, delete, and update the classes and vocabulary without need to reclassify; it must hold continually expanding digital information documents (which is most thorny but definitely need to be urgently solved).

Based on this, reasons may be given why enumerative classification used in libraries will not suffice. The largest and the foremost reason is that enumerative classification tends to classify works by a single characteristic, and basic on this multifaceted and thus be based on different aspects. It is important to have a number of classification options because this will allow for a variety of approach options for users.

In the inconsistency of technological welcomed and environmental disgusted situation, his theory was not applied into practice. Faceted classification had to wait for a breakthrough. The arrival of the Internet liberates it. With this liberation, I will argue that faceted classification is also needed from the aspect of Internet "information retrieval".

Before the advent of the Internet, although "information retrieval" becomes gradually independent from shelf organization, it used to mainly concern with searching/retrieving information from library card catalogues or scientific databases. It has its own specialty which is different from the Internet. It was mainly a closed system for scientists. For example, online retrieval system before the Internet is both expensive and limited for the general public to use, which is much different from Open Public Access Catalogue (OPAC) we now have. Then this started to change with the arrival of the Internet. Now the Internet has attracted the attention of not only the experienced users but also the common users.

The rapid growth in the quantity and complexity of information sources on the Internet means that the Internet is a unique case for "information retrieval". This specialty is especially expressed as factor of users. Users need a high level of precision with rapid response, which trumps the requirement for depth and breadth of the resources. Although information will remain the same, a researcher will require more different interpretation than that of a child. Topics can be related to authors. For example, a piece of information can be about Hemingway, who can be also related to the author (novels by Hemingway). With only the keywords, we necessarily cannot give judge of linkages and relationships among information. Therefore when users factors are took into consideration with subject, then these differences must be taken into account.

All these factors, of course, have already existed in the library environments in some way. For the limitation of "one book, one place", the users can not be that grasping. Since the Internet eliminates the limitation away, all these factors float out and need to be considered in organization of information for later easy searching/retrieval. It is also important to stress the specificity of the Internet medium, because the making of the Internet contents is often accompanied by misunderstandings that are connected to the naive copying of data from the given reality into the Internet medium when classification is mentioned in general. When library people do efforts to give change from the library environments onto the Internet environments, they never give much attention to the influence of the change on the methodology of classification.

The concept of information on the Internet environments is nothing like any of the concepts of information, increasing and varied though they may have already existed in the library environments. This increase is so sudden and rapid and the changes not only in the quantity but also the quality. Subjects, objects, applications, users, media, locations, and languages can classify these Internet resources. Different from shelf organization in the library environments, classification on the Internet is the problem of what information is about and not where physical documents (books) are in the library shelves and also different from "information retrieval" before the Internet, this information is not about what information in what kind of periodicals.

The ideal for "information retrieval" is to design systems which will allow searching to be carried out as efficiently, simply and economically as the complexity of information and grasping needs of the users. The ever-increasing complexity and specificity of the Internet information need to be matched by elasticity and minuteness of detail in organization. Classification is required to have the ability of not only classifying documents, but also people, objects, places, and events. Information has many more dimensions as argued above. If all of these dimensions are reduced to a single (one-dimension) standardized representation of that information, we necessarily lose linkages and relationships among these knowledge products. A dimension means a facet, so that the above statement means that more facets are called for Internet "information retrieval". We need classification, which will pay attention to one and at the same time all the co-equal mutually exclusive aspects of information. To provide structure to the various sources available on the Internet, I will argue in detail that how faceted classification can be said to be very relevant and useful in classifying the documents in the Internet environments. It provides multiple perspectives to manage the information. Different users may have different needs and different understandings of classification of information. The users may choose the perspective that matches their own needs. Facet classification has more chances to be a suitable choice and should be proposed, which can provide with multi-approach for information.
single characteristic it lists all subjects and provides symbols for them. They are based on the use of controlled terms that are composed by a fixed number of classes which resulted in fixed positions for subjects that have existed when a system was premeditated [Ramshirish]. So it has “a rigidly specified network of pathways leading to rigidly grouped collections of items” [Vickery, 1966].

While faceted classification differs from enumerative classification in that facets, so distinguished, are not locked into rigid, but are left free to combine with each other, so every type of relation between subjects may be expressed. As information continues to grow quantitatively and qualitatively, it is impossible to enumerate all these changes [Mills, 2004]. Faceted classification can realize these requests in practice.

Faceted classification focuses on the essential and constant characteristics/facets, which is useful for micro-grained rapidly changing information repository. It can be used to create deeper and more complex knowledge structures by exploring variants of combination. Because complex subjects are not listed, such classification schemes are easier to compile. Classification schemes are shorter for the reason. Despite brevity, they permit classification of both very simple and very complex subjects. New subjects can often be catered for by the combination of already existing concepts. Relative to enumerative classification, faceted classification makes a term that can be used more consistently both in indexing and in formulating search questions; thus, creating an extremely important method of precise retrieval, especially to the Internet in which information changes faster than the library environments. In a faceted classification scheme, if the “fundamental categories”/facets are sound, new entities can be described. These characteristics make faceted classification keep the merits in “information retrieval”.

Firstly, it has advantage in providing for the flexible combination of terms to represent a specific complex/compound subject, thus it does not have to provide lots of ready-made complex/compound classes. It lists individual facets and classifiers can join them together as they are needed. It usually includes pre-coordinated classes, but they do not need to be as numerous as in enumerative. Broughton discusses the advantages of faceted classification over enumerative classification in that faceted classification avoids having to pre-assign all compound terms to the classification; avoids arbitrary assignment of complex subject headings to a single place in a classification; subject headings can be synthesized from underlying facet elements [Broughton, 2001].

It is more flexible and suited for dynamic domains in continual expansion and change. Faceted classification allows various characteristics to coexist without prearranged relationships. It is able to accommodate new topics, whether these happen as a result of the interaction of existing subjects. In faceted classification even if you do not know the name of an object, you can achieve a very accurate shared understanding of what it is, by describing it in terms of several mutually exclusive foci/isolates from different facets [Patel 2002]. If you are trying to describe a refrigerator, for example, you can convey what it is very accurately by specifying its size, the substance it is usually made of, its color, its typical location in a house, and its primary function etc. Faceted classification groups entities by conceptual facets through analysis and not through division. Instead of a deductive process of dividing a set of entities according to application of one immutably essential characteristic at each level in the hierarchy, the set of facets does not have to be exhaustive, either during classification or retrieval by information seekers. Creators of a shared knowledgebase can add a new facet at any time. Classifiers can add a new facet at any time and users can select elements from as few or as many facets as suits them [Vickery, Foskett, 1959].

Secondly, it often incorporates much more specific detail of a subject. They are flexible, classifying from different viewpoints or facets, which does not necessarily have to be limited in achieving predetermined static classification avoiding subjective assignment of complex subjects to a single place in enumerative classification. For example, information on “forensic medicine” might be classified under medicine only, while the legal aspects are not recognized. Faceted classification would recognize both set of elements. Also it can deal with the fact that information users might seek a resource from any number of angles corresponding to its rich structure and multidimensionality. In faceted classification, it is possible to describe a dog as an animal, as a pet, as food, as a commodity, as long as the “fundamental categories”/facets have been established with which to do this. It is a more organized approach to allow fast growing information and increase potential of information discovery of different needs of user to the same piece of information in the Internet environments.

Once subjects are divided into component single-concept or single-characteristic parts, which can be used to provide subject access by combining them according to the interests and the needs of the users by combining needed foci/isolates (elements in facets) from separate facets in a structured way to express complex/compound subjects. They can provide for more flexible combinations of terms. So in one way, they can bring conveniently together all the aspects of a special field of knowledge, aspects that may be scattered in the universal enumerative classification. Also they can itemize the concepts in that field in more detail. Therefore, more research is necessary for the development of faceted classification geared toward online organization of information and later for searching/retrieval. The implementation of faceted classification on the Internet may be useful, because it provides a common language to describe and identify the contents of digital documents.

Faceted classification is very accommodating to new concepts and new composite subjects, so it is quite appropriate for the fast growing the Internet environments. It is more attentive to the users need because every user’s need is different. In that case only faceted classification can take care of this. Faceted classification allows for exploration directed by the users, where a large database is progressively filtered through the users’ various choices; until arriving
5. Advantage of Expressive Aspect of Notation in Faceted Classification

So far, I have said little about notation and it is of secondary importance, but this does not mean that we need give no thought to it. On the contrary, in the history of libraries, some otherwise useful schemes have failed to win success because their notation symbols.

The main purpose of the notation in library classification is to maintain the order of classes, since without notation there is no way of establishing an exact and sole sequence. Although the order of classes in a classification scheme may be logical, it has no natural filing order, which alphabetical order or numerical order possesses. If one user wants to know, for example, where parrots fit into the sequence, or whether physics comes before or after mathematics, each class must have an ordered symbol that indicates its position relative to other classes. We can of course scan randomly arranged collections with our eyes to identify the term that we want. But visual scanning of such a mass is however, usually too slow, perhaps we can find the term until our eyes ache. So for convenience it is necessary to arrange the terms in some conventional order, as alphabets or numbers naturally tell an order.

There are many qualities of notation and many factors to influence them. I will especially argue two factors of them, which have close relation to faceted classification. They are hospitality and expressiveness, which are mutually exclusive. In the library environments still the same physical limitation; although expressiveness has the advantage of assisting the users to following the structure of the classification, nevertheless it is achieved only by lengthening the notation, which is great damage to the hospitality of notation. Since the hospitality is more important for shelf organization, expressiveness has not been considered as a merit for notation. In case of faceted classification, notation plays a very important role in synthetic of facets, but it is criticized to be seriously over-complexity and not suitable for putting on the spine of a book, or requiring the users to memorize the whole notation symbols to locate the particular book in the libraries. Although Ranganathan worked with great care to give the shortest possible combinations for complex subjects, they often themselves are difficult to be short for the need of synthetic.

Obviously, if we anticipate this kind of flexibility, we must pay the price of long notation symbols. We can see from the example of “Treatment of Eye Diseases for Children”, which has the notation as “L 185: 4 6 9C” in Colon Classification [Ranganathan, 1960]. With the canon of Relativity, which states that the length of notation symbols is relative to the complexity of its subject, Ranganathan has demonstrated that on average his notational symbols for a given subject are shorter than those of Decimal Dewey Classification (DDC) [Ranganathan, 1957]. Ranganathan himself seems to underestimate their influence on the popularity of a classification scheme in the library environments [Foskett, 1974]. But this has changed on the Internet environments. I will further base my argument on comparison of difference of the library environments and the Internet environments.

Usually most users choose as simple a method as possible, for example, search engines. They do not want to be troubled with learning completely theoretical and complex theory of faceted classification with lengthy notation. One defense can be that this criticism is not substantial in the Internet environments, as we do not have to worry about the physical location of a document. One of important merits of the Internet environments over the library environments is that electronic information is not limited to physical locations. The Internet information classification can be liberated from notation symbols, since the problems of assembling notation symbols and putting information in one sole place disappear on the Internet. There is no need of a particular order because we are only indexing, not shelving [Pollitt, 1997].

With the liberation from the library environments, I will especially contend that expressive aspect of notation symbols in of themselves is still an essential feature of faceted classification for Internet “information retrieval”.

It can be argued that notation has potentials to permit multilingual access to information. Ranganathan expressed (faceted) classification as the process of translation of the name of a specific subject from a natural language into a classification language. Notation is the translator [Ranganathan, 1951]. Since (faceted) classification systems use notation independent from a specific language, users can enter search terms in a given language and those terms would then be related to the relevant parts of information in different languages (notation as a switching/connection) [Mai, 2004]. For example, in UDC, “599.735.32” will always used to express for “Moose”. The verbal description can not be that important to input in any languages, for the concept of Moose is always “599.735.32” in the computer stores.

The use of notation has an important advantage in the international environments like the Internet, since it can be used to represent subjects independently of natural languages. It really does not matter that much “moose” is called “oosika” or “herajika” in Japanese or “Mi Lu” in Chinese, if we agree that it can be represented by “599.735.32”. Classified information can be retrieval irrespective of language of the information since the notation symbols are independent language. In this sense, using of notation symbols essentially has a potential merit for multi-languages information communication.

So instead of the natural languages, expressive notation symbols, in which the length of symbols should reflect the position of the classes in the whole, will be especially useful in browse searching/retrieval, and also keyword searching/retrieval [Rosati, Lai, Gnoli; 2004]. The expressive quality of notation is valuable because it helps the users
to find their ways about the systematic order. The loss of expressiveness makes the overall arrangement much harder to following. Notation can have more functions than just as a natural filing order to maintain the order of classes, like knowing where women studies fit into the sequence, or whether philosophy comes before or after theology.

Successful notation is one that at a glance the order of the scheme can be seen, from an individual notation symbol for a basic element in the classification, the importance of this piece of element in relation to the whole structure. To achieve the qualifying of subjects by the combination of symbols, to show related subjects, to analyze information of several subjects, to record aspects such as forms, places and languages. Expressive notation can also help users as a map of the subjects. When users look at this expressive notation, they can see how the topic they have chosen is in the context of the structure. They can see more general classes, and more specific classes. They can control it further to narrow or broaden the topic they want to search for. In retrieval request, the users may or may not be able to state their need precisely. If they cannot, classification can provide some map of structure of the territory they are exploring, to help them narrow their searches and avoid irrelevant matters. Even if they can, such a map structure may be useful if it calls to mind the relationships with other parts of that territory.

As to keyword searching/retrieval, this can be expressed as that notation is very important in faceted classification because it is the notation that controls the order of the schedule and the building of complex/compound subjects. It can be easily broadened or narrowed by expansion or reduction of the length of numbers as the above example. Here in one way, we can further take advantage of notation to dig out sets of relevant items in a more micro-way; that is, notation of facets can show clearly the position of different facets of a complex/compound subject [Mills, 1960].

An expressive notation is appropriate to information organization in the Internet environments [Slavic and Cordeiro, 2004]. For the expressive aspect of notation, mnemonics aspect would be sacrificed. It results in long notation numbers, which is fatal to classification for shelf organization, for long notation cannot be fulfilled into the spine of a book as mentioned above, and is difficult to remember. But to the Internet users it would be less important, for it can be handled by the computers, so user will not be bothered with memorizing notation [Broughton, 2004]. While users conduct searches with terms in keyword searching/retrieval, there is no need for the user to really “see” the structure of faceted classification with expressive notation if they would not like to. Notation symbols do not necessarily be displayed in every search form: keyword(s) will be understood more directly by the most users, and notation symbols could be shown in advanced display as additional information, allowing those users who would like to understand the logic and operation of faceted classification as basic of “information retrieval”. Faceted classification brings a basis for systematic order and control to the collection so that an information package can be retrieved according to a particular aspect of its character.

Ultimately, function of Internet faceted classification is more for “information retrieval”. It is different from that of faceted classification in the library environments, which is for shelf organization. Theoretically, the disadvantage of long notation symbols can be useful as argued above. So what is urgent now is to try a faceted classification scheme suited for Internet “information retrieval”.

6. Conclusion

To conclude, in the Internet environments, faceted classification has potential to be the best choice for organizing information for later easy searching/retrieval in the “information retrieval” system. The liberation of physical limitation from the library environments, faceted classification has more chances to show its abilities, and also emphasis on multi-dimensional aspects of information and multi-level of retrieval request of the Internet users, the need of faceted classification becomes more and more urgent. Qualities and factors of notation should be reconsidered. I especially argued that expressiveness of notation should not be easily ignored since there is no worry about the physical location of a document on the Internet; more attention should be paid to its usage as a potential switching/connection to multilingual information and expressive notation as a significant point in browse searching/retrieval and keyword searching/retrieval.

In all, of course, in general it requires the coordinate cooperation in a whole organic “information retrieval” system. In this paper I have argued that faceted classification keeps the merits of classification in information retrieval comparing with search engines and overcomes the problems with a traditional enumerative classification structure by providing for the flexible combination of terms to represent a specific complex/compound subject. Especially I have argued that expressiveness of notation in faceted classification, which has been thought of as a defect until now, can find new usage as a potential switching/connection to multilingual information and a significant aiding in browse searching/retrieval and keyword searching/retrieval. As only a first attempt in a whole “information retrieval” system, much more work is necessary.

Acknowledgments. I wish to thank for those useful references; especially I wish to express my gratitude and indebtedness to my supervisors Prof. Sekimoto and Sakata for all the help and training. I wish I had attended more conscientiously to all the suggestions and comments given by them. I also express my gratitude to kindly advice of Prof. Yamazaki in Chuo University, apart from being my supervisor.
Almost the sole place to discuss faceted classification now on the Internet, although not that formal one. Details can be found at http://www.poorbuthappy.com/fcd/

A number of other terms are mentioned meanwhile when we talk about classification; thesauri, taxonomies, controlled vocabularies, ontology, etc, all of which represent a long history of structures for information. We’ve heard all of these terms in the context of the Internet. For the purpose of this paper, classification here is in the library context.

In the context of libraries, cataloging and classification are usually used in tandem with one another. Although there are obvious similarities between them, the differences between them have significant implications for the constitution of “information retrieval”. I will give a separation of them. Here, “classification” is uniquely useful for describing the content (subject matter) of documents, while cataloguing describes particular physical aspect of that document. (It includes documents title, authorship, date of publication, unique identity, ISBN, URL, etc., length of document, etc.) of course, there is also the view of catalogue as that it describes subjective aspect (what is physical document about, which is the work of classification) and also particular physical aspect of that document. Therefore, catalogue is a broader than classification and it is also called as surrogate of classification. I subsume that catalogues (metadata on the Internet) is separate from classification as the third step in the
For easy comparison of classification and retrieval technology, I use a very general term on search engine much like to be easily argued back by others. It is little wonder that for the last ten years or so there has been a flood of information retrieval systems. There are many ways to give type of search engines and the realm of concept of search engines is different too. We have been showed with new terms for new techniques and also new terms for old ones. An earlier generation called subject indexing now so-called “subject gateway” or “information gateway” are not much different when focusing on the second step of “information retrieval”, that is, they use roughly the same classification schemes to give organization of information, the difference is mainly on the fourth step in searching/retrieval in which those “subject gateway” or “information gateway” may introduce new search engine technology. Also “subject gateway” or “information gateway” and the Internet information retrieval systems (such as Google, etc.) do not have difference in quality from the fourth step of searching/retrieval technology, but difference is mainly from the second step when the way of organization is mentioned, in which those “subject gateway” or “information gateway” use classification scheme to give organization to information while search engine systems do not give organization with classification. (With reference from: B. C. Vickery, D. J. Foskett. Classification and indexing in science. Second edition. London: Butterworths, 1959 Pvi)

Classification, originally an independent intellectual exercise-first, invaded bibliography and librarianship. In order to tell difference of knowledge classification as an independent intellectual exercise, in the realm of libraries, classification is specially called as bibliographical classification. That is library classification is equal to bibliographical classification. Other opinion on bibliographical classification is that in order to give difference of classification for shelf organization. It is specially related information retrieval (called indexing in the realm of libraries). According to Foskett, D. J. “information retrieval consists of four main stages: Identifying the exact subject of the search; Locating this subject in a guide which refers the searcher to one or more documents; Locating the documents; Locating the required information in the documents” (Foskett, D. J. 1974.) Traditionally, in the days when a research worker accumulated his results until they would fill a book, so searching is based on unit of “book”. In the libraries, in order to find one piece of information, finally one must first find the carrier of that piece of information, namely the book. Organization of information in the libraries is finally related to organization of books on the shelves. It is not at once obvious to tell difference of information retrieval and shelf organization when classification is mentioned. Therefore, these differences are premised to be pointless in my argument.

The need of faceted classification for information retrieval has its start with the International Study Conference on classification for information retrieval, held at Dorking in 1957 and detail can be found in the researches by CRG. But CRG’s argument has its background of library environments, known as subject index, either manual or machine-based.

Sayers has summed three qualities as brevity, simplicity and flexibility. And Mills gave the same grouping in his 1960’s book.

“It is sometimes called “hierarchical” notation, as compared with a purely “ordinal” notation” [Mills, 1960 p. 40]