DOI: 10.5958/0976-2469.2021.00020.8

LIBRARY HERALD
Vol 59 No 3
September 2021

Developing Taxonomies and Ontologies Using Ranganathan's Facet Analysis Approach

GARIMA GUJRAL*

The paradigm of facet analysis developed by Dr. S.R. Ranganathan has been the backbone of futuristic approaches to build knowledge organization tools and taxonomies. Disparate pieces of information in the form of metadata have the potential to form and depict meaningful relationships as useful sources of information. Organizing and structuring disparate pieces of information is primarily done by knowledge managers, taxonomists and ontologists across organizations. Ranganathan gave library science its own vocabulary, "facet", "isolate", "phase", "focus" that are used for technical services. Facet analysis is a process that involves analysis of a subject or a domain into its facets based on a set of postulates, canons and principles. This article throws light on the application of facet analysis for developing knowledge organization tools such as contemporary e-commerce product taxonomies and ontologies. Facet analysis is designed on the core idea of the natural way of thinking and understanding information. This study traces the relevance and importance of Ranganathan's approaches of classification as a basis in developing taxonomies and ontologies.

Keywords*: Knowledge Organization, Taxonomy, Facet Analysis, Dr S.R. Ranganathan, Information Organization, Ontology, Semantic Web*

0 INTRODUCTION

Knowledge Society is growing at a skyrocketing pace. The growth and development of knowledge requires a systematic approach to organize knowledge. Logical principles guide for the purpose of knowledge organization. Dr S.R. Ranganathan's faceted approach has influenced the development of several futuristic and modern-day classification schemas. Classification initially used in library and information systems to provide guidance to the users and provide them relevant information resources in a time bound manner has applications in the information technology domain. It is widely used in complex information technology based environments such as the semantic web.

* Taxonomist, Walmart Global Tech, Bengaluru, India.

The analytico-synthetic approach of knowledge organization used till date in multidisciplinary domains was developed by Ranganathan. He reflected upon the ways knowledge is embodied in published materials and how users would seek them. Information retrieval has replaced the traditional reference services and is backed by the ability to synthesize text and qualitative data in a variety of ways. His facet analysis approach has been compared to a source of light for guiding travelers and saving them from taking wrong turns. As quoted by Ranganathan "Then there will not merely be the goal be glimmering, but every inch of the way will be illuminated." [1] The formula of facets and phases of a well-designed analytico-synthetic scheme of classification will serve the purpose of such a light". The traditional bibliographic classification systems lacked the expression of compound subjects. Facet analysis is not restricted to library and information science but has also been the core for designing several information retrieval thesauri. The theory has been presented as a series of 46 canons, 13 postulated and 22 principles.[2]

Ranganathan elucidates upon classification as an uncovering of the thought content using concepts such as facet, phase, and analytico-synthetic approach.

1 REVIEW OF LITERATURE

Ghosh and Panigrahi (2015)[3]applied Ranganathan's analytico- synthetic approach to develop a domain ontology in library and information science. They analyzed the domain hierarchies and developed the taxonomy using Protégé ontology editor which proved the importance of Ranganathan's philosophies in developing ontologies and knowledge organization tools.

Gopinath (1992)[4], presented the basis of Ranganathan's theory of facet analysis and identified its application for the analysis of unstructured knowledge. He discussed its value in knowledge representation in various contexts.

Spiteri, [5]explored Ranganathan's theory of facet analysis and presented it as a simplified model that is based on the principles of facet analysis by Ranganathan and Classification Resource Group. The main purpose of the model is to enable to understand and apply it in developing cutting edge technology based systems.

2 APPLICATION OF ANALYTICO SYNTHETIC APPROACH IN KNOWLEDGE ORGANIZATION SYSTEMS

The growth and development of the semantic web requires it to be well organized in a structured manner. Both conceptual factors and computational efficiencies are a must while handling complex knowledge organization systems

such as conceptual taxonomies and semantic webs. While traditional forms of taxonomies were mostly hierarchical in nature, more and more faceted taxonomies are also being used to organize knowledge. Concepts in the form of relationships are used to frame disparate sets of data into meaningful information. Taxonomies are evolutionary in nature and have the ability to assimilate new structures. Metadata is the core of any knowledge organization system. Classification systems could be in the form of verbal indexing systems or a more structured system in the form of classes and sub-classes.

The analytico-synthetic approach is used to denote concepts belonging to the universe of knowledge in which a subject is analyzed as facets initially in the idea plane and synthesized in the verbal plane followed by the notational plane. He postulated Five Fundamental Categories. As per the postulate each isolate facet should be broken down as one of the Five Fundamental Categories i.e., Personality, Matter, Energy, Space and Time (PMEST).

3 BUILDING AND MAINTAINING TAXONOMIES

The following steps need to be undertaken to design and maintain taxonomies based on the analytico-synthetic approach[1]-

i. Understanding the scope and domain of ontology
ii. Identifying organizational and user needs.
iii. Identifying the nature of data in hand.
iv. Understanding the scope and limitations of data.
v. Identifying the domain's classes, subclasses, and properties
vi. Segregating main classes from the attribute values.
vii. Developing classes and sub-classes hierarchy.
viii. Determining the facet values and properties.
ix. Pilot testing of the taxonomy tree.
x. User feedback and experiences.
xi. Improving the taxonomy tree based on user feedback.
xii. Maintaining and documenting the taxonomy.
xiii. Defining the taxonomy governance standards and methods.

Developing a taxonomy for a particular domain requires a specific viewpoint. The viewpoint helps us to understand how users would conceptualize and visualize a phenomenon. The purpose of developing a taxonomy is to ensure better browsing and retrieval. The domain has been represented accurately so, all the concepts are highlighted and have a source node.

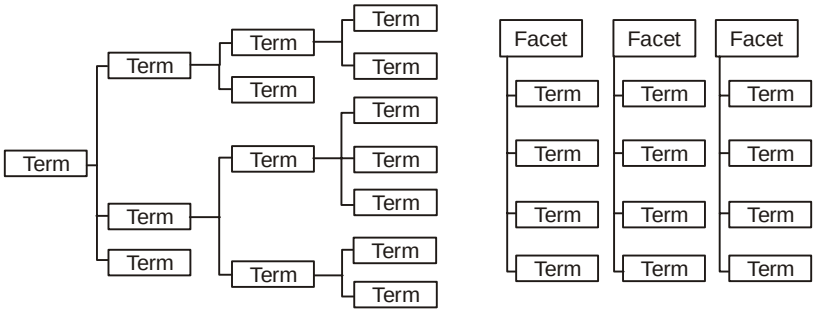

Figure 1: Hierarchical v/s Faceted Taxonomies

Figure 1, represents the hierarchical v/s faceted structures of taxonomies, hierarchical taxonomies comprise terms to be organized in a parent child relationship. Faceted taxonomies exists as flat lists and also have a relationship across domains. While facets exist as describing characteristics and are highly granular in nature such as size, colour, battery type, screen size to name a few in case of e-commerce product taxonomies. A top-down hierarchical approach as well as a faceted approach in developing a taxonomy can be applied. A taxonomy with the combination of both is also gaining popularity. [8]

31 FACETED APPROACH FOR TAXONOMY DESIGN

The information search and retrieval methods have completely changed over the decades and users seek more flexible ways to access information. The faceted approach provides users with ease and greater flexibility.[6]A facet could be mutually interdependent or dependent on other facets or attributes. The hierarchical approach can be used to organize and browse through data. Mutually independent facets describe only a single aspect of information. As per the faceted approach each concept should denote a single concept. Two processes need to be used for designing a taxonomy: analyzing the metadata and synthesizing it into new classes. Analysis involves controlling the forms and reviewing if the concept would be added as a class or an attribute.[9] The answers are not universal and depend on the kind of taxonomy that needs to be designed.

4 DESIGNING FACETED TAXONOMIES

Faceted taxonomies are replacing the traditional hierarchical taxonomies. Faceted approach allows better integration with search and also serves as filter and also provides a better user experience for both end users as well as

subject matter experts. Facets can be customized and a single set of facet may not be suitable for multiple kinds of user needs. It entirely depends on the subject areas and the used cases like enterprise or e-commerce taxonomy for implementation. User interaction with the content in an effective manner can be achieved through a faceted taxonomy. Facets can define the multiple attributes and breakdown the entire content set size of the metadata and bibliographic data.

The nature of facets varies for e-commerce taxonomies such as product or item specific attributes, e.g., enterprise based taxonomies can have facets such as departments, functions, roles etc. Facets for an e-commerce taxonomy depends on the kind of products and that could be size, colour, fabric, length and can be customized based on the needs.

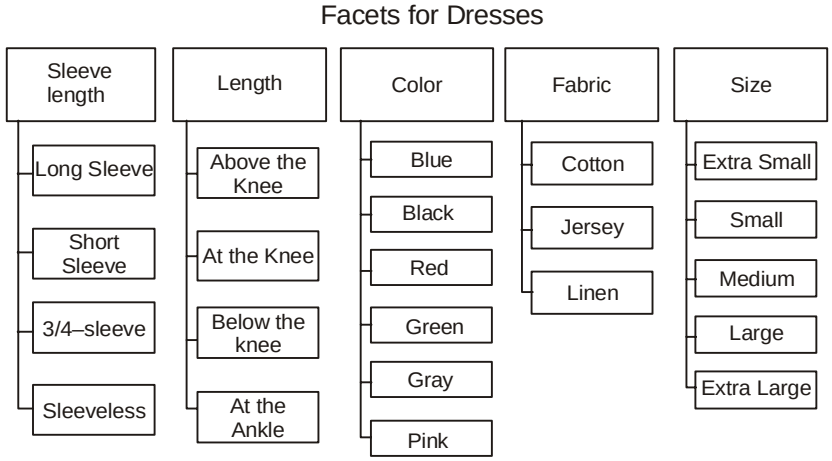

Figure 2: Faceted Taxonomy for E-Commerce

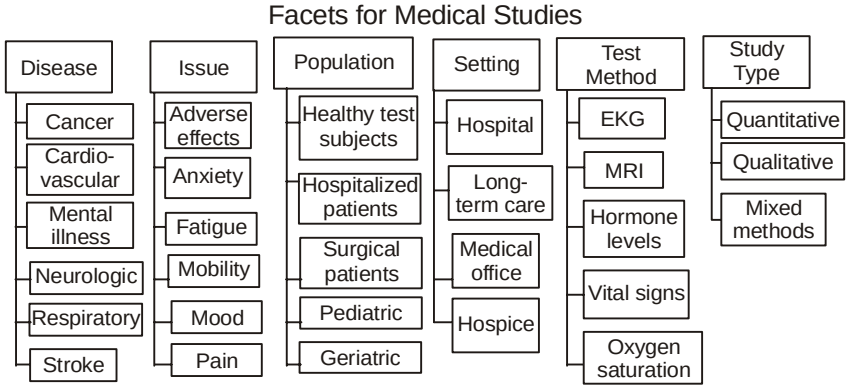

Figure 3: Facets for a Domain Taxonomy

Figure 2, illustrates an example of a faceted e-commerce taxonomy and its value. Taking dresses as an examples the various kinds of facets are length, colour, fabric, size, material. These facets define various fundamental characteristics for an item and can also be use as filters for the purpose of retrieving suitable search results based on user needs.

Figure 3, illustrates the facets for a domain specific taxonomy. Taking an example of the medical science field the facets can be organized and categorized on the basis of disease, issues, populations, settings, testing methods or types of studies. If we align the above listed domains to Ranganathan's analytico-synthetic approach the facets are derived on the basis of the five fundamental categories- Personality, Matter, Energy, Space and Time.

Facets can also be derived from the narrowest term in a hierarchy depending on the level of hierarchy. They can get specific and detailed, the levels of depth should not be ignored and derived from each layer of the data.

For example: hierarchies designed as a reciprocal inverse relationship of Broader Term (BT) and Narrower Term (NT)-

- Soccer is a kind of sport
  Soccer **BT** Sports;
  Sports **NT**Soccer;
  Soccer has broader concept Sports;
  Sports has a narrower concept **Soccer;**
- Operation Theatres are in Hospitals.
  Operation Theatres **BT** Hospitals;
  Hospitals **NT** Operation Theatres
  Operation Theatres has broader concept Hospitals
  Hospitals have narrower concept **Operation Theatres**

The examples above represent the decision making process for deriving out facets at certain levels of hierarchies.

The content should be analysed at first before designing the facets and attributes. A faceted taxonomy needs to be tied closely to the content. It is an essential part of the knowledge management and information retrieval process. Faceted taxonomies also provide a well-defined structure of controlled vocabularies with several dimensions and to query and increase the retrival.

5 TERMS FOR A FACETED TAXONOMY

The primary purpose of a faceted taxonomy is to allow users to limit and filter out the search results. It provides enhanced results for the end users. While designing the facets the terms used for the purpose of content acquisition

must align with creating subject based facets and tagging the areas. The terms used to describe facets can depend on the subject or the product category in case of e-commerce taxonomies. The subject term is essential for the tagging and retrieving the content. The terms may comprise a whole set of hierarchy. The terms may be named keeping in mind a combination of both top down and bottom up approach and looking at several content items and touch points. The terms are assigned on the basis of the descriptive indexing.

For example: A taxonomical hierarchy for several domains such as web application, decision making processes and globalization changes has been enlisted as follows:

BT (Broader Term): Health Informatics

NT: (Narrower Terms):

–Dental Informatics

–Nursing Informatics

–Bioinformatics

–E-Health

–Centralized Information Systems

–Health Care

BT (Broader Term): Human Resource Management

NT: (Narrower Term):

–Labour Laws

–Recruitment

–Unionization

–Employee Benefits

–Payroll Processing

BT (Broader Term): Gender

NT: (Narrower Term):

–Cultural differences

– Identity

– Gender & Society

Every narrow terms are sub domains within the broader terms and can be used interchangeably beyond the contexts of broader terms.

The transition should take place from a narrower term to the related term. Narrower terms can be refined in order to be used as a broader term. Hierarchical relationships can be built in an independent manner and universally.

The terms need to be reviewed and merged into a single concept and should be tested through an iterative concept. The terms define trends, patterns and subcategories emerge out of from the terms. The terms can be derived out of broader as well as narrower terms.

6 DESIGNING ONTOLOGIES ACROSS THE SEMANTIC WEB

The World Wide Web has revolutionized the way we connect and interact with people as well as our information access patterns. Cutting edge technologies such as artificial intelligence, machine learning models, linguistics are taking over the web. These approaches based on meaningful relationships are called as the semantic web. The concept of semantic web was first given by the World Wide Web Consortium (W3C), by Tim Berners-Lee.

Ontologies are a knowledge organization tool used by the semantic web. Ontology can be understood as defining something as how it exists and how it is described. Ontology is multidisciplinary in nature which falls under the field of information science and computer science. It provides a fixed set of vocabulary representing concepts and their relationships. It allows searching through unstructured metadata in the web architecture in an efficient manner.

While taxonomies are hierarchical in nature and comprises related terms and have associative relations. Ontologies have a custom created semantic relationships. Term attributes may not be a primary requirement for taxonomies but for ontologies.

For example: a taxonomy may have generic relationships-

Healthcare Industry **RT** (related term) Hospitals and Hospitals **RT** (related term) Public Health Sector

In an ontology, we may have customized, semantic relationships for healthcare industries comprising of hospitals aiding to the public health sector.

Customized relationships for every domain may not be entirely semantic in nature. But hold a parent-child or hierarchical relationships. Semantic relationships helps in identifying directional relationships across domains of knowledge.

7 CHARACTERISTIC BASED FACETS FOR TAXONOMIES AND ONTOLOGIES

Faceted taxonomies are designed in the forms of attributes, filters, and dimensions to limit various search results within the deep semantic web. Several factors should be considered while deciding the kinds of facets:

- Nature of the content.

- End user interface design.
- Facets in the form of filters or attribute values.
- Purpose of the knowledge organization tool.
- Kind of values to be added.
- Tagging process support.
- Level of expertise for the end users as well as taxonomists.
- Keyword ranking and purpose.
- Business needs and best practices.

8 INFORMATION SEARCH AND RETRIEVAL BASED ON RANGANATHAN'S PRINCIPLES

Ranganathan's approaches continues to be the basis and provides for achieving the following aspects in contemporary information architecture–

- Information search and retrieval
- Alternate search patterns
- Information recall and precision
- Narrowing search results
- Broadening search results through interrelated domains
- Better access to contents and information resources
- Eliminating search errors and no matches.
- Creation of term lists, subject headings, and classification schemas.
- Defining relationships and connections between subjects and attributes. e.g., Semantic networks, ontologies, thesauri etc.

9 CUSTOMIZED SEMANTIC RELATIONSHIPS USING RANGANATHAN'S ANALYTICO-SYNTHETIC APPROACH

Ontologies are a complex form of taxonomy that consists of richer information. They comprise associative relationships, related terms and custom created semantic relationships. Term classes are an essential component of the taxonomies defined on guidelines based on web ontology language and resource description framework. The key purpose of ontology is to describe and illustrate the domain of knowledge and it supports indexing, categorization, tagging and retrieval. The semantic nature of the ontologies is an integral element in knowledge organization structures.

An ontology can be considered a type of taxonomy with even more complex relationships than in a thesaurus, which the following graphic represents.

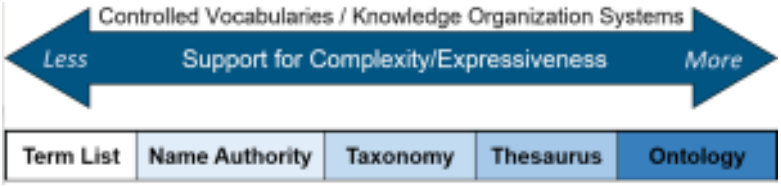

Figure 4: Knowledge Organization Systems

In other words, a controlled vocabulary that has the features of semantic relationships, classes of concepts, and attributes for concepts, can be considered a kind of ontology, but there are other definitions and understanding of ontology within the field of information/knowledge management.

While we usually refer to "controlled vocabularies" as the over-arching category for these things, it is probably better to go up a further level and call an ontology a kind of "knowledge organization system," rather than a kind of controlled vocabulary. Controlled vocabularies are kinds of knowledge organization systems, where the emphasis is on managed terms or concepts for the purpose of tagging or categorizing and information retrieval.[9]Ontologies, by themselves, are not necessarily for information retrieval, at least not directly. And this is one of the points of differing definitions of ontologies.

10 CONCLUSION

Developing knowledge organization tools like taxonomies and ontologies requires specialized training and approaches. Collaboration among information science professionals and domain experts is required to understand the existing semantic mapping and design knowledge organization tools.

Understanding, applying Ranganathan's contributions and approaches in developing contemporary knowledge organization tools in a well-defined and structured manner is a must. Building ontologies and taxonomies using his principles have been proven to be great in terms of maintenance and governance of taxonomy hierarchies. Application of faceted approach can be seen for defining semantic relationships in designing enterprise, domain as well as e-commerce product taxonomies. Customized semantic relationships can be carried out as per the five fundamental categories. The difference between hierarchical and faceted taxonomies helps information science professionals in designing the best information retrieval systems as per user needs. Maintaining consistency for large taxonomies is important to come up with governance standards. The terminologies need to align with user search keywords. Nowadays folksonomies i.e. user defined tags and taxonomies are also becoming

a key consideration in designing taxonomies and those terms can provide greater insight to information professionals. Integrating fundamental principles with cutting edge technology based approaches can help leverage the benefits of the knowledge society.

REFERENCES

1. RANGANATHAN (S R). Prolegomena to library classification. Ed 3, Asia Publication House, New Delhi, 1969.
2. RANGANATHAN (SR). Colon Classification, Gopinath M A, (Rev and Ed), 7th edn, (Sarada Ranganathan Endowment for Library Science; Bangalore), 1987.
3. GOSHA (S) and PANIGRAHI (P). Ranganathan's Analytico-Synthetic Approach in Developing a Domain Ontology in Library and Information Science. *Annals of Library and Information Studies*, 62, 1; 2015; 274-280.
4. GOPINATH (MA). Ranganathan's Theory of Facet Analysis and Knowledge Representation. *DESIDOC Bulletin of Information Technology.* 12, 5; 1992; 16-20.
5. SPITERI (L).A Simplified Model for Facet Analysis. *Canadian Journal of Information and Library Science.* 23, 1; 1998; 1-30.
6. GARFIELD (E). Father of Library Science in India: A Tribute to S.R. Ranganathan. *Herald of Library and Information Science.* 62,1; 1995; 151-163.
7. KATE (W). S.R. Ranganathan: A Chronology. 1892. London Library Association.
8. HEDDEN (H). Controlled Vocabularies, thesauri and taxonomies. *Indexer.* 26, 1; 2008; 34-40.
9. HEDDEN (H). Taxonomies and Controlled Vocabularies Best Practices for Metadata. *Journal of Digital Asset Management.* 6, 5; 2010; 279-284.
10. HEDDEN (H). Taxonomies and the Information User. *Information Outlook.* 14, 8; 2010; 10-13.