

Notice of Violation of IEEE Publication Principles

"Classification Model based on Ranganathan's Facets Theory for Information Organization and Management"

by Zhigang Ji, Qingyu Sun, and Gang Liu

in the 2010 International Conference on E-Health Networking, Digital Ecosystems and Technologies (EDT 2010), 2010, pp. 413 – 416.

After careful and considered review of the content and authorship of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles.

This paper contains significant portions of original text from the paper cited below. The original text was copied with insufficient attribution (including appropriate references to the original author(s) and/or paper title) and without permission.

Due to the nature of this violation, reasonable effort should be made to remove all past references to this paper, and future references should be made to the following article:

"Proposition of a Faceted Classification Model to Support Corporate Information Organization and Digital Records Management"

by Sabine Mas and Yves Marleau

in the 42nd Hawaii International Conference on System Sciences 2009 (HICSS 2009), 2009, 1 – 10.

Classification Model based on Ranganathan's Facets Theory for Information Organization and Management

Zhigang Ji¹ Qingyu Sun² Gang Liu³

¹The department of the Library, Hebei University of Engineering, Handan, China

²The department of the Library, Hehai University, Nanjing, China

³The department of the Library, Nanjing Forestry University, Nanjing, China

Abstract—The facet classification practices currently taking place inside many public organizations demonstrate the flexibility, expressiveness and simplicity of the faceted classification scheme. Set against the thematic content of digital documents (nearest to individual needs) and the context in which documents are created (inseparable from record management needs), faceted classification offers many advantages. The goal of such an approach is to suggest a classification language that is unified, referential and stable through time, consisting of a limited number of predefined facets.

Keywords— *Information Organization; Classification; Information Management; Ranganathan*

I. INTRODUCTION

Despite recent technological advances, most large organizations still face significant information management problems. These problems are often related to the means of describing and organizing digital documents in order to enhance find ability, share ability, usability and lifecycle management.

To respond to the need of managing, finding and using digital documents, records managers have advocated for the use of the already available paper-based corporate classification scheme [10] [12]. With this tool, based on a hierarchical structure of descriptive categories related to functions and activities of the organization, it is possible to physically regroup all documents pertaining to the same activity or project. It is recognized that the corporate classification scheme helps organize documents in the immediate or long term and facilitates the search and retrieval of documents by all employees [12].

II. DOCUMENTARY CLASSIFICATION

Documentary classification refers to the organization of documents by class based attributes or common division criteria. The first common division criteria is also the most important as it is this criteria that is used to describe and organize information holdings in logical classes. However, even the most simple documents have many attributes (e.g., type of document, topic) that can be used to group them under various classes. The number of possible groupings can be infinite and the same documents can be classified in different ways.

Since the advent of micro-computers, users have been using a hierarchical interface to organize administrative digital documents. This hierarchical interface has cognitive advantages for the individual as it enables him to locate a document through local browsing of a tree-structure directory representing the spatial location where the document is classified [3] [4]. However, inherent limits to such a structure, already noted by many researchers [2] [13], compel users to choose a classification based on personal criteria that will sometimes reflect either the document type, the document topic or the activity the document is attached to, where the document in actual fact could be classified in many places simultaneously.

III. PRELIMINARY IDENTIFICATIONS OF FACETS APPLIED TO DIGITAL DOCUMENTS ORGANIZATION

A. Facets theory

Traditional institutional hierarchical classification schemes are enumerative schemes. They have to include all existing subjects within a given domain and anticipate the emergence of new subjects.

Excerpt from a classification scheme:

- Human Resources Appraisal
 - Regular employees appraisal
 - Contract employees appraisal
 - Casual employees appraisal
- Human Resources Recruitment
 - Regular employees recruitment
 - Contract employees recruitment
 - Casual employees recruitment

In such a classification scheme, repetition of certain concepts such as the different categories of employees that are the object of Appraisals and Recruitment is inevitable and necessary.

A facet, by contrast, is "a part of a subject". Faceted classification can be seen as a combination of complementary conceptual groupings rather than a long structured list of subjects. The concept of faceted classification applied to a document's subject was introduced by S.R. Ranganathan. Ranganathan's classification scheme, and the sufficiency of this faceted classification scheme when applied to digital documents in a business context.

Compared to a simple hierarchical and unidimensional documents organization mode, faceted classification systems are much more efficient as they have the capacity to integrate

descriptions reflecting various dimensions of the information object. This permits an easier access to information by providing multiple navigation paths to any document.

For example, in the excerpt above we can see that the Human Resources process is composed of at least two different concepts: Appraisal or Recruitment (Action) and Employees (Object of the action). We could decompose the previous example into two distinct facets: Action and Object.

Action	Object
Appraisal	Regular employees
Recruitment	Casual employees

Concepts represented by terms, are enumerated within each facet, independently from each other, and combined, on demand, in a clause for a given situation. When saving a document, the employee can choose between one or many pre-established values within one or many facets with the help of a drop down menu.

B. Identification of mandatory facets

A preliminary exploration of the possible usage of facets applied to digital document organization was conducted within the Government of Quebec in an effort to develop an integrated paper and digital document classification system. After the completion of this project, we tested the feasibility of using faceted classification to organize personal documents of a University of Montreal employee. With these two studies we were able to identify essential facets that should be part of a faceted classification system. We were also able to observe the influence of the user's position on the selection of candidate facets for the model. The results of these two research projects will be discussed in some detail in this section in order to provide support for the proposed facet model in section 4.

An initial study conducted in 2001 for the Workgroup on classification and indexation of the Quebec Government's documents came to the conclusion that the hierarchy of traditional classification systems had to be "broken down" and that one of the most efficient ways of doing this was to use a set of facets. We tried to develop a more detailed and rigorous facet analysis approach for the "Business Process" and "Type of document" facets. These facets had already been identified as potential facets by the Workgroup. The main objective of our analysis was to better characterize the nature of the suggested facets and clarify their definition. We also verified that the set of facets covered all elements of the corporate classification scheme. For example, we had to preserve the context of creation of the information object (e.g., identify the administrative activity to which the document is related).

IV. FACETED CLASSIFICATION MODEL SPECIFICATIONS

A. Model requirements

Results from these preliminary studies motivated the creation of a faceted classification model that could support corporate information organization and digital records management. To be functional, this faceted classification model, described below, has to meet various theoretical, methodological and practical demands. On a theoretical level,

the model has to reflect a set of classification norms and principles related to document management (e.g., preservation of the development context and the creation of a faceted classification (relevance, permanence, exclusivity, etc.). From a methodological standpoint, the plan must be constructed according to rigorous methods and procedures that can respond to the problematic and specific record management needs of each organization (e.g., business analysis). From a practical standpoint, the plan must be able to sustain the functionalities of decision-making, research and information retrieval for an efficient organization and reliable records management.

B. Model components

The faceted classification model created to organize the digital documents for large organizations that consists of three basic elements: context facets, content facets and a "hybrid" facet (a facet that contains at the same time context and content elements as a file case or "dossier"). Context facets provide a classification that is related to organizational functions and activities, for example, whereas content facets are focused on classification of the document's subject. The hybrid facet connects these two aspects of the document categorization. Each of these elements will be described in detail below.

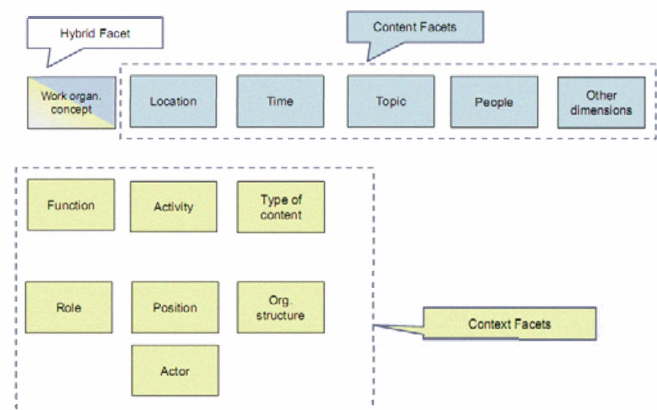


Figure 1. The faceted model components.

1) *Context facets.* Context facets describe the links between the content of a document and the administration or operational background that are the functions and management production activities.

The suggested context facets are: 1) FUNCTION, 2) ACTIVITY, 3) CONTENT TYPE, 4) ROLE, 5) POSITION, 6) ORGANIZATIONAL STRUCTURE and 7) ACTOR.

A digital document is created within a specific context. Digital documents originate from a voluntary intention of publishing [1] or are created within the framework of a specific function or activity. Context allows the reader to determine the sense of the document and makes it possible to avoid reading or interpretation mistakes [8]. This notion of context is fundamental in an administrative or business setting where the digital document is not only a collection of data or information but is rather defined as the initiation or a

consequence of a business process or activity. In this particular case, the context allows the document to be considered as a specific entity- a witness of an activity. In order to eyewitness an activity, the document must not only "encompass" the content but also the structure and the context [11]. In library and information sciences, the principle of preserving the background or the origin of the documents when regrouping documents and files based on organizational functions and activities allows the administrative context of the document to be maintained, ensures a better interpretation and authenticity, and illustrates the value of the information contained within these documents [8].

2) *Content facets*. The content of a document can be divided into several parts: title, author, recipient (if applicable), date, reference to the information context (program, process, decision, etc.), the body text and operational or transaction data, as well as complementary information such as attachments or illustrations. The above mentioned elements describe what the document is about or what has been said or done with that document. These elements constitute the memory of the action and are added by the user in the order of their importance. Part of the content of a digital document can be extracted automatically (e.g., author or date of creation) depending on the structuring level of the document. However, other parts are analyzed "manually" (e.g., the body of the text of message).

Ranganathan's facet theory offers a possible syntax for the expression and "manual" representation of a subject. Ranganathan argues that the subject of a document touches at five elements of reality: PERSONALITY, MATTER (substance, property, quality), ENERGY (main action, the action we are describing as it relates to the information object), SPACE and TIME. In a business context, a subject classification scheme based strictly and only on these five elements would have the benefit of being comprehensive, but it would also be overly general. As we have mentioned before, a facet must be researched in a pragmatic way by examining the documents pertinent to the domain and by conducting conceptual analysis that can convert a theoretically unlimited number of terms to a limited number of fundamental categories. Consequently, content facets adapted directly from Ranganathan's 5 elements, (e.g., TIME, LOCATION, TOPIC, PEOPLE) are not sufficient to describe the organizational context, which demands more precision on thematic content. Furthermore, it would be necessary to add other dimensions to these facets to comply with organizational needs such as language of the content (particularly important for federal governments and international organizations), or the entity or agent (individual or organization) implied in the transaction.

In the proposed model, content facets are identified and selected using a document method analysis. This ascending method analysis similar to the index process (analysis of the content and translation of the document language) allows identifying content facets (such as place, entity, theme) used by the multidimensional representation and indexation of the

subject of a given document in function of the particular organizational needs.

3) *Hybrid facet*. This type of facet is useful, for example, when a new project, committee, program or budget is created. The concept related to work organization is used as a medium to automatically derive, when classifying the document, content facet values (e.g., TIME, LOCATION, TOPIC), associated with this project, committee, program, budget, etc.

4) *Semantic relationships between facets values*. To facilitate and accelerate the description and classification process of the documents used by the employees, facet values that are part of the scheme are linked semantically with each other. By establishing these relationships, the model is built dynamically as each of the facet values is selected. Furthermore, the management of the life-cycle of the documents becomes more transparent.

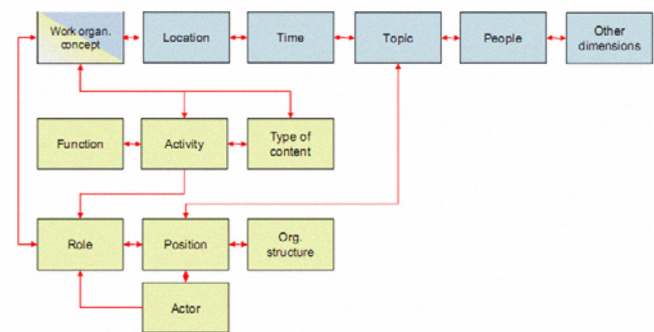


Figure 2. Semantic relationships between facets values.

For instance, by knowing the position of an employee (e.g., internship coordinator) other values could be automatically derived from the login information located in the system. That is, when an actor logs in his name (e.g., Joe Doe), other facet values such as role (e.g. coordinator, administration), topic (e.g., plagiarism) and activity (e.g., administration of internships, program evaluations), relevant to that particular position, as well as the content types (e.g., Student file) associated with these activities, will be calculated.

C. Foundation for records management services

The flexibility and the dynamism of a faceted classification scheme allow the integration of other functionalities related to the record management and life-cycle of a digital document.

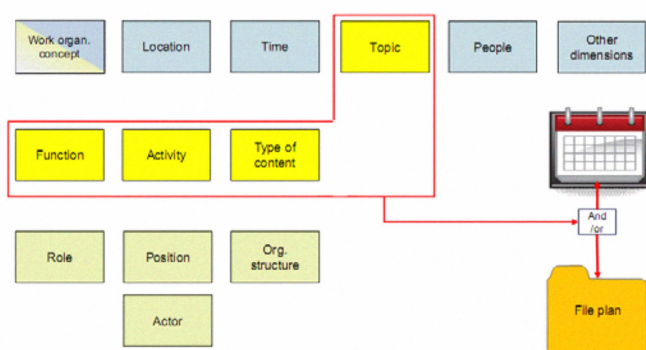


Figure 3. Links between the faceted model and records management tools.

The evaluation of documents and the disposition rules contained in the records schedules are usually put in place in relation to the creation context, subject. Since the classification and the logging are associated, once these values are selected in the faceted plan, it will allow associating the index of classification of the organizational plan with the correspondent classification rule so as to determine the retention period and the disposition mode of a document.

The proposed faceted classification improves the description of the content of the informational object as well as the context in which it was created. By establishing semantic relations between the facet values, the faceted classification scheme supports a much larger automation process for the classification and organization of information.

V. DISCUSSION AND FUTURE WORK

Facet classification:

- Allows users of the classification system to overcome difficulties encountered when using hierarchical one-dimensional organizational classification schemes.
- Improves research accuracy and management of organizational information through its life-cycle.
- Creates a larger automation, organization and classification process with the intention of freeing the user from information management tasks that rarely get done well.

The use of faceted classification for the organization and management of organizational document legacy will require a novel methodology. For this methodology to work, it is necessary to combine:

- Methods inspired by facet theory, library sciences and record management. This allows fulfilling, at the same time, long-term management of information, from both an organizational and legacy point of view, through an accurate content document analysis that draws on an administrative and operational perspective.
- Methods for the creation and development of certain content facets. These are based on a top-down approach that allows the active participation of users in the definition of the basis and maintenance of the knowledge management system.

This language allows for a multidimensional classification as well as a consensual and standardized set of documents at the core of the organization, performed by actors belonging to the same line of work. A standardized methodology that can instantiate this faceted model within an organizational context is currently being developed. This will be combined with the Contextra application to allow for the creation of organization-specific faceted classification systems.

REFERENCES

- [1] Bachimont, B., "L'archive numérique: entre authenticité et interprétabilité", *Archives (Revue de l'Association des archivistes du Québec)*, 2000-2001, 32, no. 1, pp. 3-15.
- [2] Barreau, D. K., "Context as a factor in personal information management systems", *Journal of the American Society for Information Science*, 1995, 46, no. 5, pp. 327-39.
- [3] Barreau, D. K. and B. A. Nardi, "Finding and reminding: File organization from the desktop", *SIGCHI Bulletin*, 1995, 27, no. 3, pp. 39-43.
- [4] Berchtold, S., A. Biliris, and E. Panagos, "SaveMe: A system for archiving electronic documents using messaging groupware", *Proceedings of the international joint conference on Work activities coordination and collaboration*, San Francisco, California, United States, 1999, pp. 167-76.
- [5] Boardman, R., "Category overlaps between hierarchies in user workspace", *Proceedings of Interact'01, the Eighth IFIP TC 13 International Conference on Human-Computer Interaction*, Tokyo, Japan, 2001.
- [6] Boardman, R., and M. A. Sasse, "Stuff goes into the computer and doesn't come out: A cross-tool study of personal information management", *Proceedings of the 2004 Conference on Human Factors in Computing Systems, CHI 2004*, Vienna, Austria, April 24 - 29, 2004. New York, NY, ACM, 2004, pp. 583-590.
- [7] Bowker, G. C. and S. L., *Sorting things out: Classification and its consequences*, Cambridge, MA, MIT Press, 2000.
- [8] Chabin, M.-A., "La cinquième dimension de l'archive numérique", *Document numérique*, 1997, 1, no. 2, pp. 205-216.
- [9] Conférence des recteurs et des principaux des universités du Québec (CRÉPUQ), *La gestion des archives informatiques*, Sainte-Foy, QC, Presses de l'Université du Québec, 1994.
- [10] Conférence des recteurs et des principaux des universités du Québec (CRÉPUQ), *La gestion des documents numériques des établissements universitaires du Québec: état de situation et planification stratégique*, 2004. <http://www.crepucq.qc.ca/documents/arch/Rapport-GDN.htm>
- [11] Conseil international des archives, *Guide pour la gestion archivistique des documents électroniques*, Paris, Conseil international des archives, 1997.
- [12] Direction des Archives de France, *Les archives électroniques : manuel pratique*, Paris, Direction des Archives de France, 2002.
- [13] Dourish, P., W. K. Edwards, A. LaMarca, J. Lamping, K. Petersen, M. Salisbury, D. B. Terry, and J. Thornton, "Extending document management systems with user-specific active properties", *ACM Transaction on Information Systems*, 2000, 18, no. 2, pp. 140-170.
- [14] Duranti, L., *Diplomatics: New uses for an old science*, Lanham, Scarecrow Press, 1998.
- [15] Fallis, D. and K. Mathiesen, "Consistency rules for classification schemes (or how to organize your beanie babies)", *International ISKO Conference: Dynamism and stability in knowledge organization proceedings of the sixth International ISKO Conference*, 10-13 July 2000, Toronto, Canada, Wurzberg, Ergon Verlag, 2000, pp. 339-344.