David M. Pimentel. 2011. Towards a typology of classificatory change. In Smiraglia, Richard P., ed. *Proceedings from North American Symposium on Knowledge Organization*, Vol. 3. Toronto, Canada, pp. 139-150.

David M. Pimentel (pimentel@syr.edu)
Syracuse University, Syracuse, NY USA

Towards a Typology of Classificatory Change

Abstract: Classifications of all types invariably change in response to shifting conditions in the information environment. Revising the contents of subject-based schemes is an important type of change, but the phenomenon of classificatory change has multiple interrelated aspects that go beyond content. Conceptually isolating these aspects offers a starting point for describing and comparing different types of classificatory change. The typology proposed here attempts to situate classification schemes within a context of use, interacting with other elements of the information environment. As the digital information landscape continues to evolve, there are increased opportunities for classificatory innovation. While hyperlinks have become a pervasive element in the repertory of knowledge organization, the hypertext technique of transclusion has received considerably less attention. Transclusion offers an alternative way to envision the relationship between digital resources and classification schemes. Examples from the English-language Wikipedia demonstrate how transclusion is used in the digital encyclopedia to embed modular subject-based schemes that supplement knowledge navigation and discovery.

1. Classification and classificatory change

People create classifications for many purposes and to accomplish different goals (Kwaśnik, 1989; Ellen, 2008) and numerous attempts to define classification can be found in the literature of knowledge organization (cf. Bowker & Star, 1999; Kwaśnik, 1999; Soergel, 1999; Jacob, 2004; Mai, 2004; Pimentel, 2007). In its most familiar form, classification is a method to facilitate intellectual access to information. Classified access offers a systematic subject-based approach to information resources. Library information science (LIS) professionals have a long history of devising and studying classification methods, as well as providing classified access to information resources. These bibliographic classifications rest on a range of professional activities, from revising the scheme itself to analyzing the particular information resource to be classified (cf. Bliss, 1935; Staveley, McIlwaine, & McIlwaine, 1967; Langridge, 1989; Svenonius, 2000).

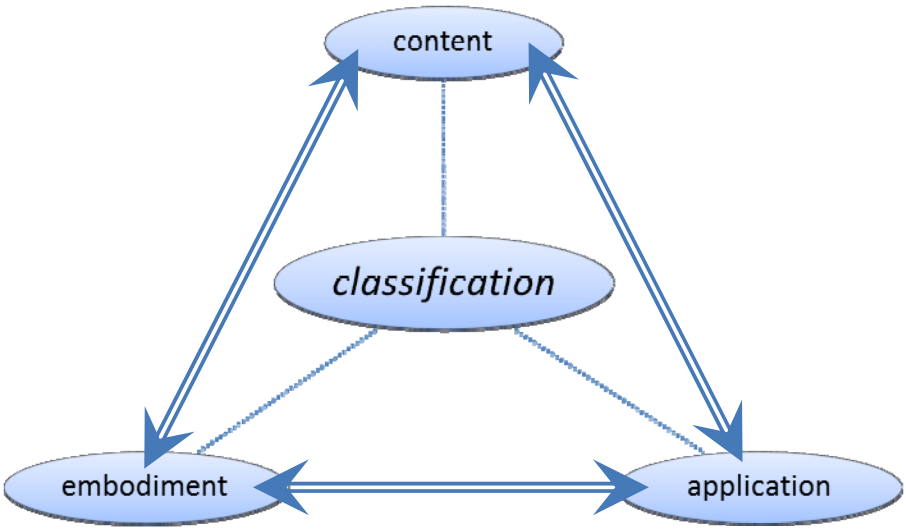

Figure 1: Interrelated aspects of classificatory change

This paper considers the concept of classification in a broad sense, effectively disregarding distinctions between formal schemes and classifications "in the wild." Classification here is understood as a dynamic process among the people, information, and technologies at play. Technologies can include print as well electronic systems, while the people involved range from scheme designers (i.e., classificationists), to classifiers, to end users. All classificatory pursuits—from design to use—are understood to occur within the context of a particular information-use environment (Taylor, 1986). The characteristics of each information-use environment dictate how its classificatory pursuits will be valued or supported.

Gaps between a classification and its information-use environment signal the need for classificatory change. Classificatory change can occur in multiple ways, in any aspect of the life of a classification in use. This paper explores the phenomenon of classificatory change by outlining a preliminary typology of three interrelated aspects of classification. These three aspects make distinctions between a classification's content, its embodiment, and its application. These aspects should be understood as a preliminary attempt to anchor the phenomenon of classificatory change: a model seeking refinement. Figure 1 illustrates the three interrelated aspects that may prompt change in a classification. The tripartite configuration shares some similarities to the semiotic triangle (Ogden & Richards, 1923), but the distinctions made in Figure 1 also reflect some influence from the three operative planes proposed by Ranganathan (1967): his idea plane, verbal plane, and notation plane. Each of the three aspects in Figure 1 is briefly delineated below.

1.1 Content

A classification's content is its principal semantic anchor: the subjects represented in the classification, what the classification is about. The content aspect in Figure 1 is largely

analogous to the plane of *ideas* (Ranganathan, 1967) as well as the vertex of *thoughts* which mediates between symbols and their referents (Ogden & Richards, 1923). Stated differently, the aspect of *content* in Figure 1 recapitulates a longstanding philosophical legacy of separating pure concepts from any particular conceptions thereof. (Popper's World 3 offers another example, as does Plato's realm of the Forms.) Any revision of classification content (e.g., establishing a new subject class, removing an obsolete subject class, merging classes, splitting classes, etc.) is an obvious instance of classificatory change. Yet not all content changes are created equal, and attempting to compare one change to another raises various questions for classification theory. Further discussion regarding changes to classificatory content appear later in this paper (section 3.2).

1.2 Embodiment

Treating the embodiment of a classification scheme as its own distinct aspect has a simple, logical basis. Namely, we each understand from personal experience how the same content can be conveyed via different carriers and media formats. Thus the embodiment of a classification concerns its tangible real-world indicators. This aspect encompasses all media for recording and presenting the classification, from print-on-paper classification schedules to interactive graphic displays. Embodiment of a scheme extends also to the visual styles and textual/graphical formatting of the classification, regardless of what the presentation medium may be. Highlighting specific subject classes as well as similar methods of emphasis and de-emphasis (e.g., font size, color, and alignment) can all be changed independently from the content of the scheme.

1.3 Application

The particular embodiment of a classification scheme has repercussions for how that classification can be used, in other words its *application*. Consider how the predominant bibliographic schemes of the twentieth century (DDC, LCC, UDC) operate in a typical library setting. These schemes were originally embodied solely in print volumes, and were intended to be applied primarily to the classification of print and manuscript items. Librarians consulted the classification schedules to identify the most specific subject class for a given resource; once the base of the classification notation was ascertained, the notation could be extended ("built") to reflect additional qualities of the item's aboutness. This process yielded a shelf mark (or "call number") that was unique to the local collection; the shelf mark was then embodied on the resource itself (e.g. spine label) as well as on the surrogate(s) representing the resource. In this way the application aspect of traditional bibliographic classifications typically served two roles. First, by translating natural language concepts into the formal language of the classification notation, it provided an indication of what the resource was about. Second, the item-specific notation served as a necessary locating device, which also facilitated access to other items on nearby topics. In simple terms the application of a bibliographic classification centered on the *what* of the resource and the *where* of the resource.

The diagram in Figure 1 shows a two-way link between the content and application aspects of a classification scheme. This reflects practical considerations of having the classification's contents in synch with its intended purposes. When a classification is needed to accomplish

certain specific user goals, this often has ramifications for content. User requirements may evolve to necessitate more robust synonym control, term disambiguation, or navigable thesaural relationships (BT, NT, RT). Evolving needs in the information-use environment drive changes in both the makeup of classificatory content as well as classificatory embodiment.

1.4 Complementary functions

The three aspects of classificatory change illustrated in Figure 1 serve as a framework to understand the dynamics at the core of classification scheme management and revision, yet they are not limited in this regard. These aspects can similarly serve the decision-making process of information professionals who must develop a rationale to employ one particular classification scheme over another. Hence one classification scheme may be chosen over another because its extant contents are better suited to the resources being classified. Or else one classification scheme might be chosen over another because its embodiment in RDF/XML is well suited to a collection of born-digital resources.

2.0 Digital contexts, digital embodiments

Inexorably changing conditions in the world around us necessitate changes in our knowledge organization systems, if they are to remain useful. Subject-based classifications in particular must have their contents regularly monitored and revised in order to adequately reflect our dynamic, shared understanding of the information environment. Revising the contents of a classification scheme for currency will always be necessary, but this is arguably insufficient without some consideration for the scheme's other aspects of embodiment and application. This paper suggests that knowledge organization theory and praxis are best served by accounting for all the ways that classification schemes can change. Furthermore, significant changes in the nature of the information-use environment should prompt the exploration of alternative means of classificatory embodiment and application.

The rise of digital libraries (specifically) and the World Wide Web (more broadly) has transformed our ability to communicate, store, and retrieve information. Yet seemingly our practices for resource classification have been slower to adapt. The increased prominence of global networked technologies in our daily lives has triggered a broad reconsideration of how we relate to information (McArthur, 1986; Birkerts, 1994; Levy, 2001; Gleick, 2011). A similar transformation regarding the development and application of classification schemes would seem possible, and perhaps even desirable. It is important here to remember that the digital environment is more than a mere shift in the storage and retrieval of resources. Hypertext/hypermedia[1] systems open radically new realms for engaging with information (Englebart, 1963; Nelson, 1965; Nelson, 1992). The implications of this sea change, often compared to the print revolution, continue to alter expectations for information interactions and

[1] *Hypermedia* is the more accurate term for describing the contemporary World Wide Web, since *media* in this sense comprehensively subsumes text, together with audio, image, and video. Despite this distinction, the literature of information science typically uses *hypertext* as the more general term, a convention followed here.

utilization. The following section introduces yet another ripple in the digital waters, the hypertext technique of transclusion.

2.1 Transclusion

Early web browsers and hypertext systems were rffrctively limited to hyperlinking between documents, with each document separate from the other. Contemporary software capabilities have evolved to permit a different type of hypertext, one that is significantly *more* non-linear. The result is a compound document that contains one or more embedded fragments; the fragments themselves remain independent from whichever documents embed them. This technique is known as transclusion (Nelson, 1992). Transclusion takes advantage of the fact that digital environments allow a single information resource to be in multiple "places" at the same time. At its core, the technique of transclusion is about reusing information resources in a modular way. Transclusion preserves the independence of all the resources involved, which allows for a change in one fragment to be immediately reflected across all its associated (embedded) instantiations. A basic example is provided in Figure 2.

Transclusion can be understood as a special type of hyperlink: instead of pointing to another resource for readers to *go to*, the link pointer brings the resource *into* the current document. Figure 2 shows document B transcluded into documents A, P, and Q. This means that documents A, P, and Q each contain a reference to document B. As a transcluded fragment, document B appears at whatever point it is referenced in A, P, or Q. In the upper row of Figure 2, the entirety of document B is the text "foo." In the lower row of Figure 2, the text of document B has been

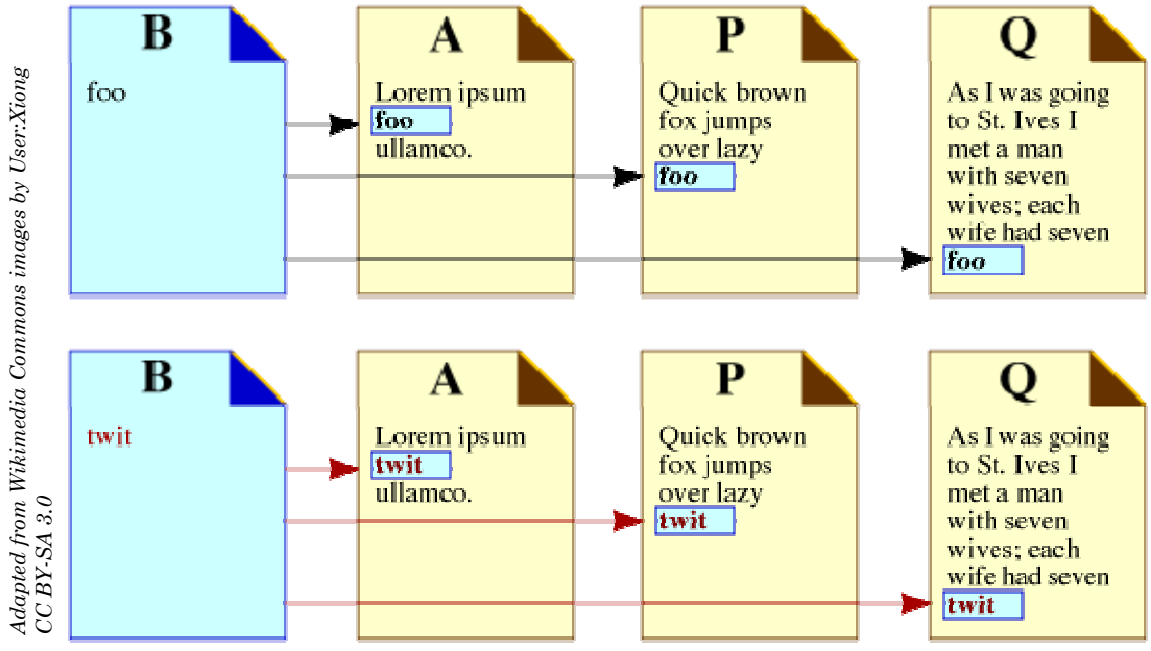

Figure 2: Transclusion, illustrated by a change in document B (before and after)

changed to "twit." As a result of the change in document B, documents A, P, and Q also change.

2.2 Uses of transclusion

The collaborative multilingual encyclopedia project Wikipedia (Lih, 2009; Reagle, 2010) employs transclusion as a way to duplicate content across its millions of articles. The English-language Wikipedia accomplishes this by using multiple namespaces (Wikipedia, 2011). The primary namespace is reserved for articles themselves, but other namespaces exist for contributors to discuss article content, for users to maintain their own profile page, as well as for images and audio files. Wikipedia has a dedicated namespace, called the Template namespace, for content that is intended to be transcluded elsewhere. Contributors can transclude a Template into any article by placing a reference (pointer) in the text of that article. Once an article includes a reference to a Template, changes to the Template will propagate to all the articles referencing it. Readers of the article are thus always presented with the most recent version of the Template content. (Returning to Figure 2, using the terminology of Wikipedia, document B is a Template that is transcluded into articles A, P, and Q.)

One common type of Wikipedia Template is the navigation template. Navigation templates are typically a classified collection of links to related content. These navigational templates are niche classification schemes, typically reflecting only one or two hierarchical levels. A single navigational template can be embedded into many Wikipedia articles, and a single Wikipedia article can embed many navigational templates. Since the navigational templates are stored as independent information resources in the Template namespace, each has its respective history of revisions. Figure 3 shows the Wikipedia navigation template for Punctuation marks, which includes links to articles together with illustrative examples of each punctuation mark. At the end of April 2011 this template provided access to 92 different articles in the English-language Wikipedia. The 92 links are classified into four groups: an unmarked (general) group, followed by Word dividers, General typography, and Uncommon typography.

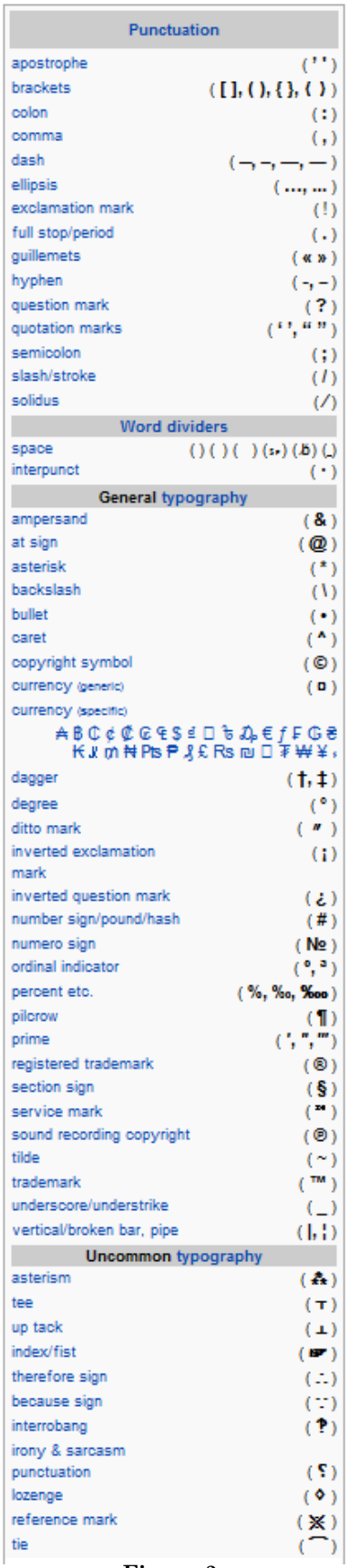

Figure 3:
Sample navigation template

The overall intent here is clear: readers who are interested in the apostrophe may also be interested in the semicolon, etc. By maintaining Templates for related content, Wikipedia contributors can quickly and easily multiply the network of links associated with any individual article by using a single reference. Similarly, as the content in that Template evolves over time, readers always have access to the most up-to-date version.

2.3 Summary

Digital environments are physically amorphous, which only increases the need to provide robust access to their information resources. Classificatory content that is embodied in hypertext offers the potential of innovative applications for providing intellectual access through the technique of transclusion. This is a substantially different approach from the method of assigning a single classification designation to an information resource. Moreover, traditionally one-shot classification notations can become outdated when the scheme content itself changes. Wikipedia's example of transcluding classified navigation templates into articles provides a glimpse into a potential paradigm shift: where classification notations never go out of synch with the classification schemes themselves, because the "notation" itself can be replaced with a window—a live, real time view—into the larger classification scheme. This section of the paper has focused on changes to the embodiment/application axis of the initial typology in Figure 1. The content of classification schemes is addressed below, together with user-based criteria to evaluate those schemes.

3.0 Users of classification schemes

Research in knowledge organization varies considerably with regard to its focus on end users. As a result, KO researchers lack shared conventions for framing analyses about people interacting with knowledge organization systems. The application aspect of classificatory change (presented above, section 1.3) explicitly acknowledges user needs and how those impact a classification's content and form. Yet classificatory change of any kind should be recognized in terms of the impact on end users. (In bibliographic systems some changes only affect classifiers, not end users, in which case considering the impact on classifiers is critical.) In order to enrich the initial typology of classificatory change with a more user-based perspective, this paper adopts four user criteria offered by Taylor (1986). Table 1 provides a summary.

User Criteria	Summary
Ease of Use	ordering elements for display, including alphabetizing and formatting to ease scanning
Noise Reduction	inclusion of relevant information; vocabulary control; excluding irrelevant information
Quality	trust in a system's reliable outputs over time
Adaptability	non-subject dimensions, ability to manipulate information

Table 1: Summary of four user criteria (Taylor, 1986: 50-70)

In keeping with Taylor's holistic perspective on all types of information-based work, these user criteria are purposefully generalizable to many aspects of designing information system. The four criteria were empirically derived from Taylor's studies of real-world KO praxis, making them logically well-suited benchmarks for understanding the impact of classificatory change (Pimentel, 2010).

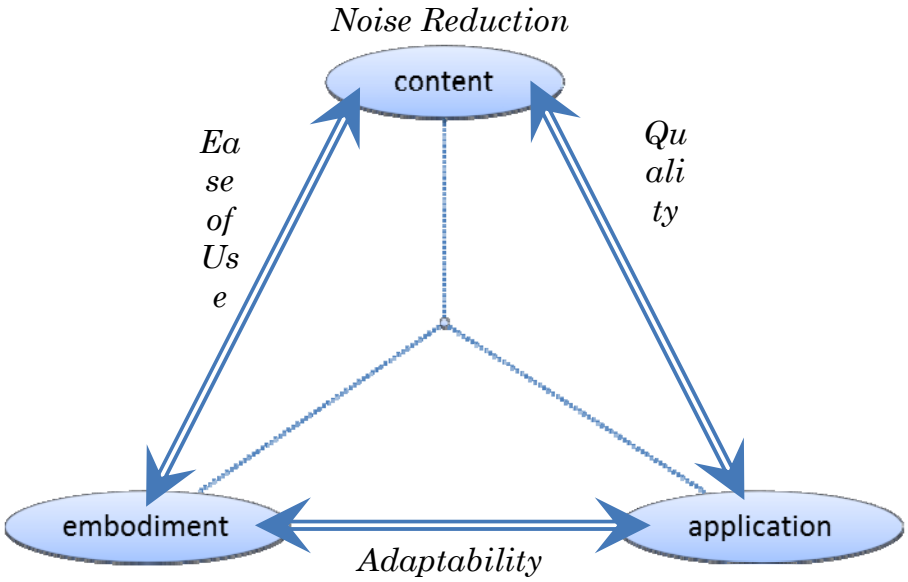

Figure 4: Classificatory change aspects correlated with user criteria

3.1 Enriched model

The criteria in Table 1 represent a high-level account of user needs and evaluative principles. Figure 4 shows the result of plotting the user criteria onto the initial model of classificatory change. This mapping indicates how the different aspects of classificatory change can directly

affect the user experience. Of the four user criteria, three were plotted along the intersections of classificatory aspects, with the remaining criteria (Noise reduction) corresponding almost exclusively with changes to classificatory content. The following section discusses changes to classificatory content as well as its ramifications for the user criteria of Ease of use, Noise reduction, and Quality.

3.2 Changing classificatory content

The job of the classificationist is rooted in coping with different types of change. Revising the contents of subject-based schemes is an ongoing endeavor in the pursuit to organize a growing body of knowledge. Here we draw a distinction between classification as an intellectual process and a classification scheme as an intellectual product. When abstracted from its embodiment and its application, classificatory content is a conceptual complex of entities that stand in relation to one another. These entities might be ideas, events, works, or indeed anything at all. Richardson (1930, 2) characterizes classificatory entities as "whatever has separate existence. Whether its substance is matter or motion or spirit is indifferent. If it is, it may be classified, and if it can be classified it must be that it is."

Since classificatory content exists in the realms of infinite creativity, memory, and possibility, the relationships themselves may be as varied as the entities—though the most common relationships represent hierarchical positions (part/whole, superordinate/subordinate), synonymy, and near-relatedness. The shifting complex of classificatory entities and relationships can be compared to the idea of a structure. Structures are an intangible notion found throughout practically any human discipline or endeavor (Pullan & Bhadeshia, 2000). In the KO literature, structure is central to the theory developed by Tennis and Jacob, for whom structure is

> …the cohesive whole or "container" created by the establishment of qualified, meaningful relationships among the components… which comprise the "bounded space" of the structure (Tennis & Jacob, 2008, 265).

These perspectives are valuable to bear in mind, since they point to the potentially intricate and thorny territory that will eventually have to be addressed by a more robust typology of classificatory change. At present, consideration will be limited to the most basic parts of the landscape. Changes to classificatory content, insofar as the end user is concerned, are summarized in Table 2.

Ease of Use	Noise Reduction	Quality
• formatting display/ appearance • highlighting important terms • logical ordering of presentation	• introducing and eliminating entities • controlling the vocabulary of entities • introducing and eliminating relationships among the entities	• currency • accuracy • reliability

Table 2: Summary of changes to classificatory content

Note that only the points falling under the rubric of Noise Reduction are strictly content-based. Other classificatory changes bleed from content into other aspects, per Figure 4.

4.0 Conclusion

This analysis has attempted to break down the various aspects of classification in order to develop an initial typology of classificatory change. Based on the premise that classifications inevitably change, this research has considered ways in which they change, exploring the phenomenon of classificatory change and considering untapped potential in our digital era. By focusing attention on the embodiment and application aspects of classification, my goal has been to advance the discussions regarding classificatory potentials in digital contexts. The hypertext technique of transclusion points to many possible applications of classificatory content.

This research suggests that classificatory change is a phenomenon worthy of closer scrutiny. In addition this research has suggested a potential framework for situating classificatory pursuits based on their effects for end users. Further study is necessary in order to develop metrics that will allow the comparison of different types of classificatory changes, together with the magnitude of those changes, across different classification schemes. Not all classificatory changes are created equal, and attempting to compare one change to another raises various questions for classification theory.

References

Birkerts, S. (1994). *The Gutenberg elegies.* Boston: Faber.

Bliss, H.E. (1935). *A system of bibliographic classification.* New York: H.W. Wilson.

Bowker, G.C. & Star, S.L. (1999). *Sorting things out: Classification and its consequences.* Cambridge: MIT Press.

Ellen, R.F. (2006). *The categorical impulse: Essays in the anthropology of classifying behaviour*. New York: Berghahn Books.

Engelbart, D.C. (1963). A conceptual framework for the augmentation of man's intellect. In: Howerton & Weeks (eds.), *The augmentation of man's intellect by machine.* Vistas in information handling, vol. 1 (London: Cleaver-Hume Press): 1-30.

Gleick, J. (2011). *The information: A history, a theory, a flood*. New York: Pantheon.

Jacob, E.K. (2004). Classification and categorization: A difference that makes a difference. *Library trends, 52* (3): 515-540.

Kwaśnik, B.H. (1989). The influence of context on classificatory behavior. Ph.D. dissertation, Rutgers: The State University of New Jersey.

Kwaśnik, B.H. (1999). The role of classification in knowledge representation and discovery. *Library trends, 48* (1): 22-47.

Langridge, D.W. (1989). *Subject analysis: Principles and procedures*. London: Bowker-Saur.

Levy, D.M. (2001). *Scrolling forward: Making sense of documents in the digital age*. New York: Arcade Publishing.

Lih, A. (2009). *The Wikipedia revolution: How a bunch of nobodies created the world's greatest encyclopedia*. New York: Hyperion.

Mai, J.-E. (2004). Classification in context: Relativity, reality, and representation. *Knowledge organization, 31* (1): 39-48.

McArthur, T. (1986). *Worlds of reference: Lexicography, learning and language from the clay tablet to the computer*. London: Cambridge University Press.

Neelameghan, A. (1971). Classification, theory of. In: A. Kent & H. Lancour (eds.), *Encyclopedia of library and information science* (New York: Marcel Dekker): 147-174).

Nelson, T.H. (1965). A file structure for the complex, the changing and the indeterminate. In: L. Winner (ed.), *Proceedings of the twentieth national conference of the Association for Computing Machinery* (New York: ACM): 84-100.

Nelson, T.H. (1992). *Literary machines: The report on, and of, Project Xanadu concerning word processing, electronic publishing, hypertext, thinkertoys, tomorrow's intellectual revolution, and certain other topics including knowledge, education and freedom (91.1)*. Sausolito, CA: Mindful Press.

Ogden, C.K. & Richards, I.A. (1923). *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism*. New York: Harcourt Brace.

Pimentel, D. M. (2007). Exploring classification as conversation. In: J.T. Tennis (ed.), *Proceedings of the North American symposium on knowledge organization* (NASKO) (Toronto, Ontario, June 14-15): 1-8. Available online: http://hdl.handle.net/10150/105879

Pimentel, D.M. (2010). Examining the KO roots of Taylor's value-added model. *Knowledge organization, 37* (1): 58-64.

Pullan, W. & Bhadeshia, H., eds. (2000). *Structure: In science and art.* Darwin College Lectures series. Cambridge: Cambridge University Press.

Ranganathan, S.R. (1967). *Prolegomena to library classification. 3rd ed*. Bombay: Asia Publishing House.

Reagle, J.M. (2010). *Good faith collaboration: The culture of Wikipedia*. Cambridge, Mass.: MIT Press.

Richardson, E.C. (1930/1964 reprint). *Classification: Theoretical and practical, together with an appendix containing an essay towards a bibliographical history of systems classification. 3rd ed*. Hamden, Conn.: Shoestring Press.

Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society of Information Science, 50* (12): 1119-1120.

Staveley, R., McIlwaine, I.C., & McIlwaine, J.H.St.J. (1967). *Introduction to subject study*. London: André Deutsch.

Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge: MIT Press.

Taylor, R.S. (1986). *Value-added processes in information systems*. Norwood, NJ: Ablex.

Tennis, J.T. & Jacob, E.K. (2008). Toward a theory of structure in information organization frameworks. In: C. Arsenault & J.T. Tennis (eds.), *Culture and identity in knowledge organization*: *Proceedings of the tenth international ISKO conference* (Montréal, August 5-8, 2008). International Society for Knowledge Organization: Advances in knowledge organization, no. 11 (Würzburg: Ergon): 262-268.

Tennis J.T., & Sutton S.A. (2008). Extending the Simple Knowledge Organization System (SKOS) for concept management in vocabulary development applications. *Journal of the American Society for Information Science and Technology, 59* (1): 25-37.

Wikipedia, the Free Encyclopedia. (2011). Wikipedia namespaces – http://en.wikipedia.org/wiki/Wikipedia:Namespace
Template: Punctuation marks – http://en.wikipedia.org/wiki/Template:Punctuation_marks