A synthetic approach to the classification of music. Review article

Rick Szostak

How to cite this article:

Szostak, Rick (2020). "A synthetic approach to the classification of music. Review article". *El profesional de la información*, v. 29, n. 1, e290105.

https://doi.org/10.3145/epi.2020.ene.05

Manuscript received on 16th November 2019
Accepted on 30th November 2019

Rick Szostak ✉
https://orcid.org/0000-0001-8570-4418

University of Alberta
Department of Economics
Tory Building, 8-19.
Edmonton, Alberta, T6G 2H4, Canada
rszostak@ualberta.ca

Abstract

This paper first reviews the advantages and disadvantages associated with both pre-coordination and post-coordination in classification. It then argues that we can have the advantages of both if we couple a post-coordinated (synthetic) approach to classification with a user interface that privileges the word order in search queries. Several other advantages of such an approach to classification and search are reviewed. It better captures the nature of a work (or object), addresses important issues with respect to social diversity, and facilitates user queries. It produces subject strings that resemble sentence fragments; this serves to clarify the meaning of terms within the subject string, and makes subject strings more comprehensible since humans typically think in sentences. These various benefits are then illustrated in the classification of works of music. It is shown that many important characteristics of works of music are best handled by such a system. These are generally poorly addressed, or not addressed at all, by existing approaches to the classification of music.

Keywords

Post-coordination; Pre-coordination; Subject strings; Classification; Music; Basic concepts; Social diversity; Grammar; Sentences; User interface; Nature of a work; Recall; Precision; Search; Information retrieval; Knowledge organization; Review article.

1. Introduction

The first part of this paper will review the latest thinking on synthetic (that is, post-coordinated) approaches to classification. It will discuss in particular how the *Basic Concepts Classification* (*BCC*) employs a synthetic approach toward the classification of documents, objects, or ideas. The second part of the paper then discusses how a synthetic approach facilitates the classification of works of and about music.

2. A synthetic approaches to classification

The third section will summarize the "traditional" view that there are both advantages and disadvantages associated with post-coordination. The fourth section will argue that it is possible to pursue a post-coordinated approach that has all of the advantages traditionally associated with pre-coordination. The fifth section then outlines several further advantages associated with this type of post-coordination. We then briefly discuss the broad outlines of the *Basic Concepts Classification* which pursues this type of post-coordination.

We should first define our terms. A "synthetic" approach to classification allows the classifier to combine different terms from a classification in generating a subject string. They might, for example, combine "subverting," and "democracy" to create "subverting democracy." Such an approach is often called "post-coordination," because the combination of terms occurs after the classifier investigates the document (or object or idea) to be classified. The alternative is "pre-coordination" in which the classifier must choose from an enumerated list of (generally complex) subject headings the heading that most closely captures the document (or idea or object) in question.

3. Pre- or post-coordination? The traditional view

Svenonius (1993), **Jacob** (2004), and **Šauperl** (2009) all agree that there are both advantages and disadvantages associated with both pre-coordination and post-coordination. They each conclude that each approach might be best suited to different contexts. **Šauperl** (2009) had wondered whether technological advances had caused one type of system to become preferable. She finds the same result as **Jacob** (2004) and as **Svenonius** (1993): each of the two systems has advantages.

The main advantages attributed to pre-coordination are speed and precision. Speed of searching, both Jacob and Šauperl argue, is better in pre-coordinated systems because complex subject headings allow for more focused retrieval, and hence for better precision. Pre-coordinated systems facilitate faster search because they are precise: The user does not get a lot of "false drops": suggestions that are a poor fit for the search query.

> We should worry that users cannot achieve a precise search until they first identify the appropriate subject heading to search for

We should worry, though, that users cannot achieve a precise search until they first identify the appropriate subject heading to search for. This is no trivial concern: Many public libraries have moved away from subject classification because of a feeling that users struggle with subject search; they instead organize their shelves the way bookstores do (**Lyttle**; **Walsh**, 2018). Even university libraries are increasingly moving away from subject searching in favor of keyword searching. This is unfortunate, for subject searching is potentially far more precise than keyword searching (**Hjørland**, 2012) –but only if users know how to perform subject searches. A post-coordinated system might have an important advantage at the "front end" of a search if users can easily combine search terms.

With respect to precision, Šauperl worries that a post-coordinated search for "philosophy of history" will find many works on "history of philosophy" that are inappropriate to the query. This argument is critical to her view that pre-coordination has advantages, for a pre-coordinated system with a heading "philosophy of history" will direct users only to works on that subject. Yet we can note here that Šauperl's conclusion depends entirely on her assumption that search interfaces follow a Boolean process, interpreting the search as "any combination of philosophy AND history." If the search interface were to somehow care about the order in which search terms were entered, then a post-coordinated search might be just as precise as a pre-coordinated search.

One further advantage of pre-coordination is that individual terms are clarified by being placed within a complex subject string: We clarify which of the many meanings of "culture" is implied when this is placed in the heading "popular culture." Yet again, a post-coordinated approach can have the same advantage if combined with an appropriate search interface, in this case one that recognizes that "popular" is an adjective for "culture."

Post-coordinated systems have various advantages. They have shorter schedules: rather than enumerate any possible complex heading, the classificationist need only enumerate a set of simpler terms that can be combined in myriad ways. Post-coordinated systems are therefore also more hospitable, as new combinations can be facilitated automatically. When the first works in Physical Chemistry appeared they could be easily classified by combining those terms; Classifiers might otherwise have to prioritize Physics or Chemistry –and miss out on the important combination of the two– until a classificationist added Physical Chemistry to the classification. Post-coordinated systems thus grow in size much more slowly. Only rarely is it necessary to add new terms to the schedules. For similar reasons, post-coordinated systems age more gracefully, for the complex strings within pre-coordinated systems gradually fail to capture how complex ideas are expressed. New ideas can be simply expressed with new combinations in a post-coordinated system.

Post-coordinated systems are easier to learn and to employ by classifiers. As noted above, they have shorter schedules. These schedules can be organized more logically, for it is easier to logically subdivide a simple concept than a complex combination of concepts. Post-coordinated classifications are thus less expensive to create and maintain.

There are also several criteria for which pre- and post-coordinated systems fare equally well. **Svenonius** (1993) claimed that they were equivalent with respect to recall, the percent of relevant documents obtained. Both pre- and post-coordination can potentially be handled by computers. Both are well suited to universality (that is, they can be applied to any subject matter), and are browsable.

Jacob, Svenonius, and Šauperl each concluded that there was no clear winner between pre- and post-coordinated systems. Each approach might prove best suited to different circumstances. Crucially, though, the disadvantages attributed to post-coordination all hinge on the nature of search interfaces. If a search interface were developed that prioritized the order in which search terms were entered, and thus could engage with the subject string as a whole, a post-coordinated approach might match or exceed a pre-coordinated approach with respect to speed, precision, and clarity.

> If the search interface were to somehow care about the order in which search terms were entered, then a post-coordinated search might be just as precise as a pre-coordinated search

4. The best of both worlds?

Interdisciplinary scholarship teaches us to always look past dichotomies and ask if there is some continuum or middle ground between alleged opposites (**Repko**; **Szostak**, 2020). We have already seen that *all* of the advantages attributed to pre-coordinated systems rest on an assumption that post-coordinated systems are associated with search interfaces limited to Boolean search algorithms. Yet it is quite feasible to develop search interfaces that do prioritize the order in which search terms are entered, and thus clearly distinguish "philosophy of history" from "history of philosophy." Such an interface could also then treat "popular culture" as a whole rather than seeking any use of both "popular" and "culture." Computer science undergraduates have design such an interface for use with the *BCC* in the past, and at present a computer scientist is developing a sophisticated thesaural interface for the *Basic Concepts Classification* that prioritized word order in search queries.

The fields of "information retrieval" and "classification" operate quite separately within information science, even though both are interested in guiding users to the information they seek. Information retrieval has focused on the design of search interfaces, while classification research has taken the search interfaces of the 1960s as given. Yet information retrieval scholars have increasingly recognized that there are limitations to keyword searching, and that there may be scope for combining retrieval techniques with some sort of subject classification. It is time to achieve greater coordination between these two fields, and work toward a combination of classification and information retrieval techniques that best guides users to the information they seek.

> "It is quite feasible to develop search interfaces that do prioritize the order in which search terms are entered, and thus clearly distinguish "philosophy of history" from "history of philosophy"

The ideal is for a user to be able to type a search query in their own words. This is what information retrieval systems allow. But keyword searching lacks precision compared to subject searching (**Hjørland**, 2012). So, the ideal then is for a search query to be translated into controlled vocabulary that guides the user precisely toward the desired information. If we can develop a search interface that translates a user query into a post-coordinated subject string, then we can combine ease-of-use and precision. Note that one of the alleged advantages of pre-coordination, speed of use, is maximized in such an approach: The user does not have to struggle to identify controlled vocabulary but is nevertheless guided precisely to the most suitable documents (or objects or ideas). If associated with an appropriate search interface, then, a post-coordinated system could outperform a pre-coordinated system in both speed and ease of use.

We are exploring the possibility that such a thesaural interface can be developed for use with the *Basic Concepts Classification* (*BCC*) using the *Universal Sentence Encoder* (*USE*: **Cer** *et al.*, 2018). *USE* has been fed very large batches of text: it identifies synonyms for words and phrases through vectors in 512-dimensional space modeled after the idea, "you shall know a word from the company it keeps" (**Firth**, 1957, p. 11). The guiding principle is that if two words are often used in a similar context they likely have a similar meaning. Happily, *USE* can potentially deal with phrases rather than just individual words. This will save classifiers and users from having to translate each word individually into controlled vocabulary. More importantly, phrases further clarify the meaning of the words they contain (as we saw with "popular culture" above). Notably, *USE* does discriminate on the basis of word order, for it places each term in the context of the phrase it is embedded within.

The entire terminology of the *BCC* (and the *Unspsc* codes used to identify goods and services within *BCC*) have been encoded into 512 floating point numbers by *USE* so that *USE* can be instructed to translate phrases into *BCC* controlled vocabulary. Users –and classifiers– can be (immediately) given ten possible *BCC* translations from which to choose. Technically, the algorithm uses a vector array built from pre-computed embeddings of *BCC* phenomena, *Unspsc* terms, relators, and combinations of relator and phenomena, and from that array it selects the ten nearest neighbors, using a mea-

Figure 1. Home page of the thesaural interface for *BCC*

sure of cosine similarity (cosine similarity is employed based on the assumption that vectors pointing in a similar direction have similar meaning).

A demonstrator version of the interface can now be seen at
https://sites.google.com/a/ualberta.ca/rick-szostak/research/basic-concepts-classification-web-version-2013/thesaural-interface-for-bcc

At present it deals best with shorter phrases. Readers can enter any phrase and be guided to appropriate *BCC* terminology (and given the *BCC* notation that goes along with that terminology). At present, they may have to search again if important terms are missing in the generated subject string. We are working on algorithms that can cope with larger phrases. We are also working on tree structures based on the hierarchies within *BCC*: the translator can then appreciate that the best place to look for controlled vocabulary for a type of painting is within the category "Art" rather than "Mathematical concepts." Note that our interface can be trained over time through repeated use (and selection by users or classifiers of particular options) to better select the best *BCC* translation of particular queries.

> If associated with an appropriate search interface, a post-coordinated system could outperform a pre-coordinated system in both speed and ease of use

With such an interface we can mimic the "front end" advantages of post-coordinated systems (classifiers and users can put together their desired subject string) while delivering the same "back end" degree of precision as pre-coordinated systems. The system allows "philosophy of history" to be clearly distinguished from "history of philosophy" without requiring the user to first ascertain what search terms are allowed within the classification. It thus combines the advantage of logical syntax with the advantage of user construction of search terms. **Šauperl** (2009) notes that users generally appreciate neither the complex subject strings of pre-coordinated systems nor the contours of Boolean searches required by post-coordinated systems. The approach outlined here spares them from both inconveniences.

5. Further advantages of a synthetic approach:

5.1. We can construct subject strings that mimic sentence fragments

The subject strings generated by the *BCC* to describe documents or objects generally resemble sentence fragments. That is, they tend to group nouns, verbs, and adjectives in the order in which they might appear in a sentence. Such subject strings may at first seem bizarre to the scholar or user familiar with existing subject headings within the *Library of Congress*, *Dewey Decimal*, or *Universal Decimal Classifications* that privilege nouns and provide terms in an order quite different from everyday speech. Yet there are huge advantages in pursuing a grammatical format. Most importantly, we spend our entire lives (once we acquire language) thinking, speaking, reading, and writing in sentences. This is the way we are accustomed to combining words. And thus we comprehend strings of words that form (part of) a sentence far better than we comprehend strings of words organized in any other way. Classificationists and classifiers may gain familiarity over time with the unusual format of subject headings, but the typical user can be expected to comprehend subject headings far less well than they understand the average sentence.

A secondary and related advantage is that a sentence serves to clarify the meaning of the words that it contains. As noted above, the phrase "popular culture," serves to constrain which of the thousands of different uses of the word "culture" we are referring to. Sentences, it should be stressed, are less ambiguous than single concepts, for the sentence provides context for all constituent concepts. This recognition of the importance of context lies at the heart of communication theory. Communication theory focuses on 'thought units' rather than concepts. 'Thought units' may be bigger or smaller than a sentence, but will comprise multiple concepts (**Gorman** *et al.*, 2010). Information scientists worry a great deal about terminological ambiguity: How can we guide users to the information they seek when the words they will use in a search are inevitably ambiguous? We should thus value the ambiguity-reducing effects of placing concepts within sentences.

The experience of *Precis* (*Preserved context indexing system*), an indexing system developed by Derek Austin and colleagues for use in the *British National Bibliography* in the 1970s, is instructive. The purpose of *Precis* was to identify a number of different subject headings with different lead entries (at a time when subject headings were still provided in card catalogues and thus it was of critical importance that a user correctly identify the lead term). Nevertheless, the designers of *Precis* found it useful to employ grammatical construction within the key elements of their subject headings. They had experimented with other types of word order before deciding that grammar was best.

> The subject strings generated by the *BCC* to describe documents or objects generally resemble sentence fragments

> "The fact that general rules of this kind can be deduced and applied in practical indexing would seem to indicate that natural language is endowed with a greater measure of underlying logic than many classificationists would allow" (**Austin**, 1974, p. 82).

It is also noteworthy that *Precis* was successfully translated into French. It would appear that the differences in grammar between the two languages were not overwhelming.

> We comprehend strings of words that form (part of) a sentence far better than we comprehend strings of words organized in any other way

Once we are comfortable with the idea of a subject string that is constructed like a sentence, then we can imagine that a classifier can easily construct such a string. They can begin by identifying a critical sentence within a document or object description. They can then simply translate that sentence into controlled vocabulary. We shall see below that in so doing the classifier best captures the nature of a work (or object).

5.2. We can in so doing pursue facet analysis without facet indicators

Facet analysis has been recommended within the field of Knowledge Organization for several decades, but has only rarely been pursued in practice. Facet analysis has two key components:

- A stress on a synthetic or post-coordinated approach to classification, in which simple terms are combined to generate a complex subject string.
- The idea that the terms synthesized will represent different "facets" of a subject. We can eschew the challenge of providing an intentional definition of the word "facet" –that is, attempting to identify the essence of the term in a couple of sentences– by instead providing *inter alia* below an extensional definition: a list of what are considered to be at least the main facets that need to be addressed.

When **Ranganathan** (1962) introduced the idea of facet analysis, he suggested five key facets. The *Bliss Classification* later expanded this to include 13 facets (*Bliss*, 2017). It turns out that each of the 13 facets identified in *Bliss* are either a specific element of a grammatical sentence, or refer to a specific schedule within the classification of phenomena within the *BCC*.

1) Thing. *Bliss* treats 'Thing' as the principal foci of interest within a discipline. But there is no need to restrict its usage in this way. *Bliss* generally treats nouns as "Things," and thus the main noun in a subject string can be interpreted as the "Thing."

2) Kind (of thing). This is the primary fashion in which nouns are disaggregated within a hierarchical classification, including *BCC*. One advantage of a synthetic approach to classification is that we can hope to pursue a logical disaggregation. We need not treat recycling as a kind of garbage simply because there is no other place to put it within an enumerated classification. Moreover, we will often be able to avoid enumerating a set of subclasses by instead achieving these through synthesizing a main class with a set of adjectives or verbs. In any case we can employ a 'kind of thing' in a sentence without needing to indicate that that is what it is.

3) Part (of a thing). As with kinds, this facet could be handled by either subdivision or synthetic construction. We must, though, take care to indicate when we disaggregate by parts rather than kinds. Since many parts can be applied to multiple things –both houses and cars have doors– it may prove superior to have a distinct schedule of parts (within the schedule of things) that can be linked to many things. This is the practice recommended for the "Material" facet below. Such a schedule would then have unique notation that would clearly indicate when this facet was being employed.

4) Property (of a thing). This is an adjective. Adjectives generally appear before nouns in English (but after in Spanish and French). In some sentences adjectives appear after a verb: Steel is strong.

5) Material (comprising a thing). Materials are a noun, but play here a role similar to an adjective. They can be captured in a sentence much like an adjective (steel girder) or following 'of' or 'made of.' Materials are a distinct schedule (with notation NM) within the *BCC*.

6) Process. This facet refers to changes within a 'thing' under study. This is a self-reflexive kind of verb: grow, decline, fluctuate, or remain stable. We need only distinguish verbs that refer to an internal process of change from the much broader set of verbs that refer to the action of one thing on another. This will obviously be the case when there is only one noun in a sentence. Within the *BCC*, these self-reflexive verbs do receive distinct notation.

7) Operation. Here we capture ways in which a phenomenon is influenced from outside. Most verb-like terms fulfil the role of operation. Thus, the use of a verb (indicated by an arrow) in a *BCC* subject string signals an operation.

8) Product. Here *Bliss* captures the results of operations on natural entities. These will take the place in a sentence of an object: a thing that is acted upon. It is odd that *Bliss* emphasizes this particular kind of object but has no general facet for "object." Grammar tells us that these are an important component of many sentences. A subject

> Sentences, it should be stressed, are less ambiguous than single concepts, for the sentence provides context for all constituent concepts

classification that did not mention "democracy" in a work about "institutions for strengthening democracy" would be seriously deficient. We should thus provide within facet analysis for a general treatment of objects. In a *BCC* subject string these will be nouns appearing after verbs.

> We do not have to employ facet indicators to signal which facet is represented by a particular notation, as is done in all existing faceted classifications

9) By-product. Same as above.

10) Patient. *Bliss* refers here to "intermediate goods:" things that are made only to be transformed into something else. Intermediate goods can be captured in a causal chain A causes B causes C. Note that in such a sentence it is abundantly clear that B is the patient and C is the product. *Bliss* assumes that A is the main 'thing': We could leave to classifiers a decision (if necessary) as to which term is most important for shelving purposes.

11) Agent. Agents act upon other things. They will generally take the place of A in the type of sentence above. Agents are generally individuals, but sometimes tools or institutions. All of these are nouns. Each type of noun carries a specific kind of notation within the *BCC*.

12) Space. Space captures any physical (including political boundaries) or spatial dimension. These are a very specific kind of noun. These are a distinct schedule within *BCC* (with notation N1).

13) Time. Time refers to any chronological or temporal characteristic: These are also a very specific kind of noun (N2 within *BCC*).

The *Integrative Levels Classification* changes the order of the above facets slightly and adds three new facets. "**Purpose or result**" generally describes a relationship and would be captured by a causal relationship: <thing> <relationship> <thing>. A "**Pattern**" is a representation, such as a poem about X. This would usually be captured by inserting 'of' or 'in' or 'about' between two nouns. "**Modality**" deals with how a phenomenon is perceived. This largely captures elements of the theories and methods and perspectives employed by scholars. These are again specific types of nouns that should have a particular place in a logical hierarchy of things. Once again classificatory notation can obviate the need for a facet indicator.

Each facet will thus be clear in context: We do not have to employ facet indicators to signal which facet is represented by a particular notation, as is done in all existing faceted classifications, including *Bliss*, *Colon*, and the *Integrative Levels Classification* (each of which is impressive in many ways). Nor does the classifier have to consciously employ facet analysis: If they have carefully selected or concocted a sentence from a document or object description that captures the most important facets of a work, they can proceed to translate this into a *BCC* subject string. They can, if they choose, check to see if they have inadvertently missed an important facet.

5.3. We can facilitate user queries

We have indicated that a classifier might move fairly directly from a sentence in a document or object description to a subject string. A user should be able to perform an analogous operation: state a query in natural language, have this translated into controlled vocabulary, and be directed to appropriate subject strings.

We shall see below that the nature of a work is best expressed in a sentence. A user query is also expressed as a sentence. At present, we interpolate between a user and the work they seek a subject heading that is not formatted as a sentence. This inevitably risks translation problems as we translate sentences into something else and then back again. A subject string in the form of a sentence can potentially guide the user fairly directly to appropriate works (or objects). We can thus combine ease of query formation with precision in search outcomes.

This all depends, though, on the user query being in –or translated into– a similar format to the subject string. And then we need to query how regular grammar is, and whether a search interface can easily translate a user query into some standard grammatical format employed by classifiers. **Szostak** (2017) surveyed the basic rules of English grammar and identified the following list of adjustments that a classifier (or computer) might make in moving from an object (or document) description to a standard grammatical format:

- Translating interrogative, imperative, and exclamatory sentences or clauses into declarative format.
- Ignoring pronouns and most determiners.
- Using only the most specific form when nouns are repetitive.
- Translating verbs into the infinitive.
- Using combinations with auxiliary verbs to capture verb tenses.
- Translating phrasal verbs and idioms into synonyms (a task for a thesaurus).
- Placing simple adjectives before nouns but post-adjectival phrases after.
- Using compound adjectival forms to capture gradation.
- Translating adjectival phrases with "that" (or similar words) into adjectival phrases using prepositions or infinitives. Ignoring or translating the rare adverb that does not appear after a verb or before an adjective or adverb.
- Using an extra set of parentheses if necessary (or some other notational device) to clarify whether a modifier is an adjective or adverb.

- Distinguishing adverbs from prepositions when the same word can be used for each.
- Ignoring the first component of a correlative conjunction.
- Addressing inverse verbs, ideally by preferring one form over its inverse.

This is a long but manageable list. And it must be stressed that most people most of the time follow a standard grammatical construction. Therefore, no adjustments are necessary most of the time. It should also be stressed that most if not all of these adjustments can be programmed into a search interface. A user employing an unusual sentence structure should still be guided to the documents or objects they seek (and note that if the search interface does not cope well with unusual sentence structure the user is still no worse off than if using standard Boolean search methods).

A final point: Some readers may immediately be reminded of struggling to learn rules of grammar in elementary school. But linguists appreciate that we all employ nouns and verbs and adjectives with great facility even if we might have struggled to identify these on a test in Grade 5. Users will not usually find it difficult to enter a grammatical search query if we do not remind them about elementary school.

5.4. We can facilitate machine searching

One of the challenges in computerization of pre-coordinated systems is that these are all characterized by ad hoc decisions difficult for a computer to master; the proposed system is resolutely logical. Since we have separate schedules of relators and properties we do not have to somehow squeeze "recycling" into a classification of things simply because there is no better place to put it. We have then further clarified the meaning of terms by placing these within subject strings that pursue a standard grammatical format. Each of the transformations outlined in the preceding subsection can potentially be programmed into a computer. The user should generally not have to worry about whether they are formulating their query in a grammatically correct format.

5.5. We can employ natural language

Svenonius (2004) worried that there is a tradeoff between employing natural language in a classification system versus applying very precise meanings to controlled vocabulary that differ from the common understandings of the terms used. As with our discussion above of pre- versus post-coordination, we should ask whether this tradeoff can be minimized in practice through a novel approach to classification. And here again it is quite possible to at least decrease the tension identified by Svenonius.

The *Basic Concepts Classification* is described below. But we can appreciate here the idea of "basic concepts" themselves. **Szostak** (2011) defined basic concepts as those which carry very similar understandings across disciplines or cultural groups. Philosophers have long debated the nature of concepts and whether it is possible for diverse individuals to have shared understandings of concepts. **Szostak** (2011) reviewed leading concept theories and argued that most if not all of these supported the idea of basic concepts. He stressed that philosophers often focused on whether precise definitions of concepts were possible; classificationists can be satisfied with a more relaxed standard: can enough similarity in understanding of a concept be achieved such that users of the classification can be guided to relevant documents, objects, or ideas? If we agree that there are such concepts –think of "chair" or "dog"– then it is quite possible to employ natural language basic concepts in a very precise manner. The *BCC* is built upon such concepts. Users should readily appreciate most terms in the *BCC*, since they are natural language terms, but most users should nevertheless have very similar understandings of what these terms mean.

The challenge identified by Svenonius really comes into play with complex concepts. Different users may then understand the terminology in quite different ways, and the classification will have to carry extensive scope notes indicating how terms are defined. "Globalization," is a complex concept that means "effects of foreign investment on employment" to some and "effects of American movies on Spanish culture" to others. But "investment," "employment," and "movies" are basic concepts for which broadly shared understandings exist. If we make complex concepts by explicitly combining basic concepts in a grammatical subject string, as in "effects of foreign investment on employment," then broadly shared understandings are maintained. Indeed, placing concepts within a sentence-like structure further enhances shared understanding, as we have seen above.

It *may be* that some complex concepts prove hard to disambiguate into basic concepts. As always, we should be careful of drawing empirical conclusions from theoretical arguments. The *BCC*, for example, lists a common set of political ideologies, even while recognizing that people might have quite different understandings of, say, "socialism." In general, though, complex concepts can be broken into sets of basic concepts. This eliminates or at least alleviates the tension identified by **Svenonius** (2004). User studies can potentially be performed on the *BCC* to measure the degree of shared understanding of *BCC* terminology.

> "Readers can enter any phrase in our demonstrator interface and be guided to appropriate *BCC* terminology (and given the *BCC* notation that goes along with that terminology)"

5.6. We can capture the nature of a work

Foskett (1996, p. 127) argued powerfully that subject headings should be co-extensive with the essence of a work. But what exactly is the nature of a work? What is it that makes a particular work unique and distinguishes it from other works? (see **Furner**, 2010). **Smiraglia** (2001) argued that the essence of a work could be identified as the ideas that the work expressed. Smiraglia's purpose was to distinguish the meaning of a "work" from the meanings of "text" and "document" and other similar concepts. To this end he surveyed not just the literature in philosophy and information science but also linguistics, sociology, and other fields. He concludes that (2001, p. 122):

> "… the intellectual dimension of a bibliographical entry is the set of composed ideas that it conveys, which is called a work."

Texts and documents can be defined in physical terms but a work is necessarily abstract: it is a set of ideas. Works thus have an "inherent nature as communicative signifying objects" (p. 54). *FRBR* (*Functional requirements of bibliographic records*) also treats a work as an abstract entity (p. 47).

Smiraglia (2001) does not define what he means by "ideas." We can usefully flesh out this term, recognizing that works either make an argument and/or provide a description. Ideas will comprise some set of: descriptions of phenomena or relationships among phenomena, causal arguments about how one phenomenon influences another, theories applied, methods applied, and perspectives applied. Classifying works along these dimensions will thus capture the nature of a work. Classifications that omit any of these (and existing classifications in widespread use omit most and handle others poorly) will quite simply not capture the nature of a work. If a work is about the effect of tougher sentences on crime rates, then the subject heading should capture precisely the combination of "tougher sentences," "effect on," and "crime rates." Note that, since ideas are generally expressed in sentences, the ideal subject string to capture the nature of a work will also generally take the form of a sentence or sentence fragment.

Most scholarly works –and perhaps most general works– are about some sort of posited causal relationship. Yet they are almost never given a subject heading that captures the essence of the causal relationship(s) that the work is about. We try and fail to classify works about causal arguments as if they were about one complex thing. Users searching for that particular causal relationship thus face unnecessary difficulty in finding it. While the particular causal relationship(s) addressed in a work are the key aspect of what a work is about, other important aspects include any theory or method or data that was employed as well as the perspective or worldview of the author. These also should be captured in our classifications. These various characteristics of a work were identified as classificatory desiderata in the *Leon Manifesto* (2007).

Importantly, **Furner** (2010) appreciates that philosophers have devoted little attention to what a work is about but much to what a sentence is about. The point to stress here is that the approach recommended here reduces the gap between "what a work is about" and "what a sentence is about" by describing works in terms of causal statements. (Note that works that describe a thing or an action rather than a causal relationship could also be captured in terms of statements about things and or effects.) We can then apply philosophical thinking regarding the nature of a sentence to understanding the nature of a work. We have seen above a general appreciation that sentences clarify meaning.

Smiraglia's analysis nevertheless identifies certain caveats. Most centrally, the ideas that comprise a work are conveyed semantically. It is thus not possible to entirely distinguish substance from style (p. 67). A work may be appreciated as much or more for the style with which certain ideas are conveyed as for the ideas themselves. Smiraglia makes special note of music, but one could reflect also on poetry or a political speech such as the *Gettysburg Address*. Yet the blending of substance and style hardly obviates the value of identifying the key ideas of a work. It does, though, indicate that some attempt to capture style would be useful. This will likely prove a much harder task than classifying substance. Some descriptors might be fairly easy to apply: humor, satire, sarcasm. Others would present a greater challenge: rhetoricians disagree about the full set of rhetorical strategies that can be employed in a work, and how each might be identified. If we can agree on a set of rhetorical strategies to place in a classification it would then be quite easy to include these in a synthetic subject string.

Smiraglia also notes that author and reader may disagree about the key ideas in a work. Deconstructionists have shown that works (especially of fiction) often contain ambiguities of which the author was not consciously aware. Yet it should be possible in most cases to identify the key causal arguments an author was trying to make (and new authors could be asked to supply such information) as well as any theories and methods explicitly applied.

Since works might be instantiated in multiple texts and documents, a decision must be made about when a work is transformed into a new work. Such a change might be primarily stylistic: a movie based on a novel is generally considered a new work no matter how closely it follows the original. But in general,

> "The degree of change in ideational and semantic content determines the point at which a text represents a new work" (p. 50).

How would we know when ideas have changed enough to declare a new work? A precise answer to such a question may never be possible. But the approach to classification pursued here would at least suggest some key questions to ask:

has the causal argument changed?; has a new theory been applied?; has a new method been applied? Positive answers to any of these would signal the creation of a new work (though we might still wonder if adding one new variable to a complex argument really generates a new work).

Hjørland (2014), after reviewing some of the confusion surrounding the term "information," suggests that information science would be well advised to understand its focus as "documents" rather than "information." Following Smiraglia, we might suggest a better focus would be 'works' (though it is also important to identify the different documents that might instantiate a particular work; we will also argue below for the advantages of classifying also objects). In any case, this section takes Hjørland's suggestion a step further, and argues effectively that we should focus primarily upon "statements." The best we can hope for is to organize statements, and also the perspectives (including theories and methods applied) from which these emerge. We might also as a field eschew that other contested term 'knowledge.' Humanity almost never 'knows' anything, but we can increase our confidence in the reliability of particular statements by compiling argument and evidence. The purpose of a classification system, then, is to guide users to relevant statements and the arguments and evidence that support these.

Floridi's well-known efforts to define "information" are broadly consonant with an emphasis on statements and perspectives. For Floridi, information must be about something, it must be meaningful, and it must be well-formed. He also stresses what he terms the "truthfulness" of information (**Furner**, 2014 summarizes Floridi's thinking on these issues). Statements of the sort we stressed above would meet the first three criteria. They would be about something and well-formed. Their meaningfulness would reflect whether they were statements about something that some user valued and came from a perspective that the user valued (see below). **Furner** (2014) argues that information science should devote much more attention than it has to "truthfulness." Our attitude toward "truthfulness" can be the same as our attitude toward "knowledge": the best that humans can do is evaluate the arguments and evidence that support any statement. We are again guided to reflect on how we can both guide users to relevant statements and facilitate their efforts to evaluate these.

5.7. We can classify documents, objects, and ideas

In recent decades, museums and art galleries around the world have developed an online presence. They often, though not always, try to provide some overview of their collections online. Users then naturally wonder if they can search across museums for particular objects. And some users naturally wonder if they can search across different types of institutions for documents (both primary and published) and museum and gallery objects relevant to a particular topic of interest. **Marty** (2014, p. 618) discusses user needs from the point of view of museums:

> "They [users] want to be able to say, "I'm writing a paper about Hercules," or "I'm researching the evolution of glass-making technologies," and find all the relevant resources in one search, in person or online, regardless of the type of collections where the records they need may be stored. They do not want to learn that most information systems are not geared toward answering these kinds of questions, and they especially do not want to discover how difficult it still is today for cultural heritage organizations to share information about their collections and enable searching across multiple institutions."

Note that museums hold some objects because they are typical of a particular time and place (say, an axe used to chop trees), but hold other objects because they are special (say, a golden axe used for ceremonial purposes). A synthetic approach allows us to classify both types of object. Museums have not found typical document classification systems very useful for their purposes, precisely because these have been designed for documents rather than objects. But a classification such as *BCC* with a detailed schedule of phenomena is well-suited to the classification of objects. The schedules of properties and relators can then serve to clarify the nature of an object (golden axe) or its purpose (for chopping wood).

We have noted above that documents can be classified in terms of the key idea(s) that they contain. It follows that we could, if we wish, classify ideas themselves. As **Gnoli** (2008) has stressed, a classification scheme should ideally be able to handle both works and ideas. And **Börner** (2006), intriguingly, envisages a not-too-distant future in which scholars no longer write stand-alone papers but rather contribute nodes or nuggets to a web of knowledge. She can be seen as operating within a long tradition in information science, from at least the work of Paul Otlet, which sought to classify both works and ideas. A classification system that classifies works in terms of ideas is clearly one way –and perhaps the best or only way– to do so. The *BCC* strives to encompass all phenomena and types of interaction, and is thus well-suited to the web of knowledge envisaged by Börner.

5.8. We can better cope with social diversity

One common criticism of existing classification systems is that they embody gender or ethnic or other stereotypes. If, for example, male nurses are treated as a special subclass of "nurse," then the classification communicates an (outdated) idea that nurses are or should be female. Many scholars of information science struggle to erase such biases from commonly used classifications. The sort of synthetic approach urged in this paper has the natural effect of eliminating bias.

> The sort of synthetic approach urged in this paper has the natural effect of eliminating bias

If subject strings are made by combination, and there are schedules of gender that include all genders, schedules of ethnicity that include all ethnic groups, and so on, then the string "male nurse" is formed in exactly the same way as the string "female nurse." Individual classifiers may still be biased in the way they form subject strings but the classification itself naturally treats all groups in the same way.

The reliance on basic concepts within the *BCC* serves further goals with respect to social diversity. It ensures that members of different groups will understand a classification in similar ways. If not for the reliance on basic concepts, an individual from one community may have trouble navigating a classification designed by a member of another community. Importantly, members of all communities will also be able to navigate the classification similarly well, as long as they possess basic literacy skills. The *BCC* thus enhances information access among disadvantaged communities. Recall that many public libraries have found that users struggle with subject classification; it is an important social goal that members of all groups should be able to access the information they need.

Basic concepts are likely also far easier to translate across languages than complex concepts. The lesser degree of ambiguity in the original language should facilitate the identification of a very similar term in other languages. Moreover, basic concepts tend to represent things that we perceive in similar ways in the world around us. We can thus potentially provide similar levels of information access across linguistic groups.

A KOS (knowledge organization system) designed by members of one community may exclude concepts deemed important by members of other communities. The hospitality of a KOS –the ability to add new terms– is thus an important consideration here. A hospitable system will be less likely to exclude subject strings of interest to particular groups. It is not always clear where to place a new term within the multi-level hierarchies of complex terms that characterize enumerated classifications. Within a synthetic approach, new terms can usually be created through a new synthesis of existing terms. When a new basic concept must be added to a KOS, this is easier in flat and logical hierarchies: One need not search multiple levels and wonder what the principles guiding the hierarchy are.

Olson (2007) famously argued that a hierarchical approach to classification may reflect a male perspective. Women may be more likely to see the world in terms of non-hierarchical relationships. A synthetic approach is grounded in a belief that authors and users should potentially be able to combine any set of concepts as they see fit. Classification systems that pursue a synthetic approach to developing subject headings still have to organize the concepts to be synthesized hierarchically, but these hierarchies can be much flatter than those within enumerated classification. The *Basic Concepts Classification*, for example, only rarely has more than three or four levels of hierarchy.

6. The *Basic Concepts Classification*

The *Basic Concepts Classification* (*BCC*; **Szostak** 2019) is a "universal" scheme: it attempts to encompass all areas of human understanding. Both general users from different cultural backgrounds and scholars from across disciplines can potentially utilize the *BCC* to find documents, objects, or ideas produced in any culture or discipline. Whereas most universal schemes are organized around scholarly disciplines, the *BCC* is instead organized around phenomena (things), the relationships that exist among phenomena, and the properties that phenomena and relators may possess. As noted above, this structure allows the *BCC* to apply facet_analysis –an approach to classification long emphasized within the field of knowledge organization– without requiring the use of facet indicators.

The main motivation for the *BCC* was a recognition that existing classifications that are organized around disciplines serve interdisciplinary scholarship poorly. We have noted above several further advantages of a classification such as *BCC*. The *BCC* may also prove advantageous for the Semantic Web, since its separate schedules of things, properties, and relators are well-suited to the "(subject)(predicate or property)(object)" structure employed on the Semantic Web (**Szostak** *et al.*, 2016, pp. 177-182).

We discussed above how *BCC* functions best in concert with a thesaural interface that allows classifiers and users to move directly from a natural language sentence toward a *BCC* subject string. We can note here that the *BCC* is also compatible with innovative visualization techniques. An interface could allow users to experiment with various sorts of changes to a search query once this has been translated into *Basic Concepts* –perhaps by simply sliding a mouse over the different elements of the query. For example, a user searching for "(dogs)(biting)(mail carriers)" could be alerted to documents addressing "(cats)(biting)(mail carriers)" or "(dogs)(licking)(mail carriers)" or "(dogs)(biting)(neighbors)". The user can thus easily follow their curiosity to a host of related subjects. This is much harder to do within pre-coordinated classifications. Note that the user can choose to alter nouns or verbs or adjectives/adverbs in their searches. The *BCC* thus instantiates a "web-of-relations" approach that allows users to find information related in a host of ways to the initial query. Users and search interfaces could likewise easily move between broader and narrower terms in the hierarchies of phenomena (or properties). They could similarly move between simple relators and more complex relators

formed via synthesis. Last but not least, the visual interface could also guide users to related material: from searching "(dogs)(bite)(mail carriers)" to "(mail carriers)(go to)(hospitals)".

The classification of phenomena expanded on a table developed in **Szostak** (2004), *https://www.isko.org/cyclo/bcc#refS*

which itself had reflected the reading of hundreds of works across all human science disciplines in **Szostak** (2003). Ten main categories of phenomena were identified: two individual-level categories of Genetic Predisposition and Individual Differences; Non-Human Environment; and seven societal-level categories, Art, Culture, Economy, Health and Population, Politics, Social Structure, and Technology and Science. These were subdivided logically in terms of "type of". Several more categories have been added to address natural science subject matter: in accord with the idea of integrative levels: These address Waves and Particles, Molecules and Atoms, Rocks, Biological Entities, Flora and Fauna, Celestial Objects, and Mathematical Concepts.

Schedules of relators were added over time, based on **Szostak** (2012), and these were given quite different notation from the schedules of phenomena. A schedule of "properties" was developed inductively over time as the *BCC* was used to classify sets of documents and objects; it also is notationally distinct. One key goal was to have very short notations so that multiple terms could be combined in subject strings. Wherever possible, notation was developed that would be easy for users and classifiers to remember.

The *BCC* was encouraged and informed by preceding efforts to develop a universal phenomenon-based classification within the field of knowledge organization (**Szostak** *et al.*, 2016, pp. 96-100, review this history). The *Classification Research Group* in the United Kingdom had discussed the principles of such a scheme in much detail over a period of many years (e.g. *Classification Research Group*, 1969). The most important influence, though, has been Claudio Gnoli, who has pioneered the *Integrative Levels Classification* (ILC, 2004; **Gnoli**, 2016; 2017a; 2017b; 2018). There are many similarities between the ILC and *BCC*: both take a faceted phenomenon-based and analytic-synthetic approach to classification. The major initial difference between the *BCC* and ILC was the development of separate schedules of relators and properties; this has led over time to an emphasis within *BCC* on the development of subject strings that follow grammatical rules (see above).

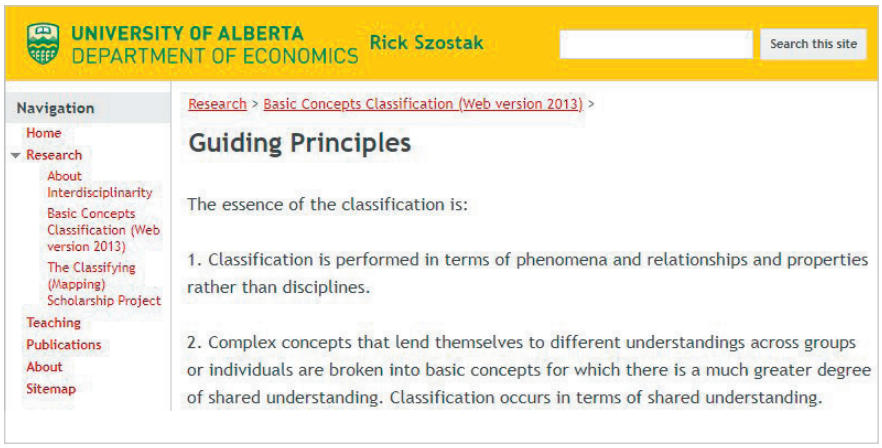

Figure 2. *BCC*. Guiding principles.

The *BCC* has been developed online at: *https://sites.google.com/a/ualberta.ca/rick-szostak/research/basic-concepts-classification-web-version-2013/guiding-principles*

Many additions have been made to the *BCC* online since it was uploaded in 2013. The schedules for the subject matter of the human sciences are largely complete (with some exceptions such as the treatment of mental disorders), but work remains to be done on several schedules addressing natural science phenomena. The schedules of relators and properties are well developed. The classification has been successfully employed in classifying dozens of randomly selected documents and museum artifacts. The *BCC* is being translated into Linked Open Data.

We have already mentioned some key guiding principles:

- Synthetic classification utilizing separate schedules of phenomena, relators, and properties
- Short logical hierarchies of phenomena
- Reliance wherever possible on *basic concepts*: those that are understood in broadly similar ways across individuals and communities
- Synthetic subject strings generally resemble sentences or sentence fragments. The use of common grammatical format facilitates search: a user's search query can be translated into the most relevant subject string.

Some other key guiding principles can be mentioned:

- For relators, several dozen key relators are developed within a handful of very flat hierarchies. These can be combined synthetically with each other and with phenomena or properties to generate thousands of very precise relators.

- Hierarchies of phenomena generally reflect a "type of" disaggregation (rarely "parts of") which reflects an ontological understanding of the world, supplemented as necessary by literary warrant, ensuring that all relevant concepts are captured.
- The notations attached to concepts are generally both short and expressive. It is thus possible to synthesize several terms and still have a notation for a subject string of manageable length. A user familiar with the classification may be able to recognize the string from the notation.
- Detailed classifications of methods and theory types are included in the classification of things so that works can be precisely classified in terms of the theories and methods applied as well. Scholars often care not just (or primarily) about what a work is about but what theories and methods were applied. It is at present generally impossible to search by theory or method applied.

7. Classifying music

Music has proven to be particularly difficult to classify. In part, this is simply because there are so many types of music. But in part it is because music is so multifaceted: it is used for many purposes, has multiple effects on its audience, conveys a diversity of themes, and reflects a wide variety of societal influences. The classifier of a particular piece of music may think it important to capture why the work was written or the effect it had (say, for a song used by a revolutionary movement). Certainly, it is easy to imagine a user wanting to search for works of music along such dimensions.

A dedicated music classification will struggle to adequately classify all of the possible influences on or effects of a piece of music –precisely because a wide variety of social, political, cultural, economic and other phenomena may be implicated. Arguably, then, the best way to capture the myriad influences on and effects of music is within a general (i.e. universal) classification that pursues a synthetic approach. Then, the classifier can have access to the widest possible array of influences and effects. The *BCC*, with its separate schedule of relators, makes it especially easy to capture the diverse influences on or effects of a piece of music.

"The best way to capture the myriad influences on and effects of music is within a general (i.e. universal) classification that pursues a synthetic approach"

By synthesizing across the schedules of the *BCC* we can potentially capture many important dimensions of works of music which are ignored in existing approaches to music classification. However, a subject string embracing all dimensions would be very long. Classifiers might then wish to emphasize only some dimensions for particular works. Typical *BCC* practice is to be guided by descriptions of works in choosing which dimensions to stress. A library that used subject strings for shelving purposes could choose which element of the subject string to prioritize (and then BOLD this in the subject string).

The synthetic nature of *BCC* means that many dimensions of music can be captured by synthesizing across diverse schedules of *BCC*. The next section reviews those dimensions for which this is surely the case, and the succeeding section reviews those dimensions for which this is likely the case.

8. Dimensions that can surely be addressed by synthesis

Existing schedules within *BCC* allow us to address synthetically several dimensions that might be important in classifying particular works. In all cases, a synthetic approach allows us to be (more) exhaustive in our treatment whereas existing enumerated schemes inevitably provide limited options which tend to be biased toward particular musical genres:

- *Intention of composer.* If desired and known, a classifier might add a motive: (to)(encourage)(nationalism). At present, it is generally not possible to indicate composer intention within subject headings.
- *Time period and place.* Classificationists have moved away over time from using terms such as "rococo" toward more objective time periods. The *BCC* allows time periods of any sort to be designated (schedule N2). [We could potentially have an Index which advised users what time period they might search for "rococo" etc.] Likewise, the schedule N allows detailed indication of geographical areas. Note that the classifier might wish to indicate when and where a work was created, but at times might instead or also indicate where it was most famously performed.
- *Creators.* Cutter numbers can be used to indicate particular composers or musicians. It should be stressed that a synthetic approach allows the classifier to potentially capture `diverse elements of creation: (Composer X)(creates)(symphony); (Composer Y)(adapts)(symphony); (Conductor X)(directs)(symphony); (Singer X)(sings); (Guitarist X)(plays). Note that by using (and) we could indicate multiple singers or players or combinations of composer, conductor etc. For choruses, or other joint efforts it is also possible to indicate gender, age, and other characteristics of a group, by synthesis. One particularly useful type of synthesis here would be when a work originally performed for one set of instruments or voices is re-arranged for a different ensemble. [Similarly, a synthetic approach allows us to capture multiple instantiations of a work, such as a recording accompanied by sheet music, or a piece created for an orchestra played by a quintet.]

- *"Kinds" of music.* (for)(wedding)(for)(ballet). The *Library of Congress Classification* (*LCC*) allows for music for children, dance music, chance music, electronic music, national music, and music of a special character. Why should these particular kinds of music be privileged? Why shouldn't the classifier be able to indicate any kind of music? Why shouldn't a user be facilitated in finding particular kinds of music? The *BCC* allows any potential kind of music to be indicated.
- *Subject.* Not all works of music have a subject. But for those which do this can be an important component of a subject heading. A subject heading for "The 1812 Symphony" should surely capture the fact that this is a piece of music about war. The subject heading for a love song should say something about love. And we might usefully get more detailed, distinguishing songs about mutual love from songs about unrequited love. As with any other work classified in *BCC*, the subject of a piece of music can be captured by synthesizing across the schedules of *BCC*. We should likely agree on a standard word such as "representing" which would introduce the subject within a subject string.
- *Effect.* The subject of a work of music may signal its likely effect: A love song is arguably intended to make us feel happy or sad, and reflect on the importance of love in our lives. The kind of music also tells us something about effect: Dance music should encourage us to dance. Yet at times it may be useful to directly indicate the effect, or at least intended effect, of a piece of music. Scholars of music still debate the effects that music in particular, and art more generally, can have on us. But we still might at times feel comfortable to say that a particular song encourages happiness, sadness, reflection, nationalism, anger or numerous other emotions or attitudes.
- *Types of analysis.* Though most of our attention in this section is devoted to works "of" music, we would also want to classify works "about" music. In general, works about music can be treated just as works about any subject –except when they analyze dimensions of music that need to be fleshed out. Music theories can be classified like any theories in schedule TT. (Music)(appreciation) is captured synthetically as is (psychology)(of)(music). So also are works about the construction of instruments. [These examples are the only types of analysis given special attention by the *LCC*. Again, the obvious question is why we would wish to privilege a subset of types of analysis. Why not treat sociology of music or politics of music just as we treat psychology of music?]

It is useful to reflect on users for a minute. What sort of searches might users wish to perform? Can we easily imagine that users might wish to search for intention or subject or kind of music? If so, can we easily imagine that different users might wish to search for quite different intentions or subjects or kinds of music? If we answer both questions positively, then we should seek a classification that facilitates such queries. This is most feasible within a general classification designed for post-coordination. It is, sadly, infeasible within all existing music classifications.

8.1. Dimensions that can likely be addressed by synthesis

With some small adjustments to existing *BCC* schedules we can likely capture several additional dimensions:

- *Traditions*. Popular, folk, and classical are all descriptors that should exist within schedule Q (of properties, or adjectives/adverbs). We could make sure that all "traditions" are captured somewhere. Note that schedule C of the *BCC* lists dozens of different cultural values. The *BCC* also allows for any ethnic group or country to be identified. It is thus potentially feasible at present to indicate any possible cultural tradition by referencing a group, place, or core value.
- *Sacred music*. Existing schemes do a limited job of indicating which parts or types of religious service a particular piece of music is intended for. We need to flesh out schedule CR on religion so that we can then link (for)(baptism) or (for)(Christmas)(service). One question is to what degree we can identify common types of ceremony across religions.
- *Techniques.* Individual techniques like "breathing" can be captured synthetically. TM codes for methods in general.
- *Intended audience*. The *BCC* has many descriptors of types of people and groups. We can thus easily signal that a particular work was designed with the French army or Chinese farmers in mind. We need, nevertheless, to identify any audience types peculiar to music.
- *Culture*. This in practice seems to mean country, ethnicity, language, and/or time period, all captured in detail in *BCC*. The *BCC* also provides a lengthy schedule CV of cultural values. Existing classifications of music are vague about what might be captured under "culture"; since the *BCC* devotes the entire schedule C to unpacking culture into subclasses, the classifier can be very precise about the cultural influences on or of a work.

Here again, we are aspiring to an exhaustive treatment, whereas existing systems classify only some traditions, types of religious ceremony, techniques, and audience. The option to associate pieces of music with cultural values is novel. Once again, a synthetic and general classification such as *BCC* allows us to serve a far wider range of user queries than can existing approaches to music classification.

9. Concluding remarks

Music is a particularly fertile ground for synthetic classification. Yet a synthetic approach allows superior classification of any set of documents, objects, or ideas. If we couple a synthetic classification to an appropriate user interface we can facilitate the efforts of both classifiers and users. We need, though, to break down the barriers between the fields of classification and information retrieval, and embrace the idea that subject strings should look like sentences.

10. References

Austin, Derek W. (1974). *Precis: a manual of concept analysis and subject indexing*. London: Council of the British National Bibliography. ISBN: 0900220422

Bliss Bibliographic Association (2017). *The Bliss Bibliographic Classification: using the scheme*.
http://www.blissclassification.org.uk/bcclass.shtml

Börner, Katy (2006). "Semantic association networks: Using semantic web technology to improve scholarly knowledge and expertise management". In: V. Geroimenko & C. Chen (eds.), *Visualizing the semantic web: XML-based internet and information visualization* (pp. 183-198). London: Springer. ISBN: 978 1 4471 3737 5

Cer, Daniel; **Yang, Yinfei**; **Kong, Sheng-yi**; **Hua, Nan**; **Limtiaco, Nicole**; S**t. John, Rhomni**; **Constant, Noah**; **Guajardo-Céspedes, Mario**; **Yuan, Steve**; **Tar, Chris**; **Sung, Yun-Hsuan**; **Strope, Brian**; **Kurzweil, Ray** (2018). *Universal sentence encoder*, arXiv:1803.11175v2 [cs.CL] 12 Apr 2018.
https://arxiv.org/pdf/1803.11175.pdf

Firth, John R. (1957). *Papers in linguistics 1934-1951.* London: Oxford University Press. ISBN: 978 0 197135488

Foskett, Antony-Charles (1996). *The subject approach to information.* London: Library Association. ISBN: 978 1 856040488

Furner, Jonathan (2010). "Philosophy and information studies". *Annual review of information science and technology*, v. 44, n. 1, pp. 161-200.
https://doi.org/10.1002/aris.2010.1440440111

Furner, Jonathan (2014). "Information without information studies". In: F. Ibekwe-SanJuan & T. M. Dousa (eds.), *Theories of information, communication and knowledge: A multidisciplinary approach* (pp. 143-179). New York: Springer. ISBN: 978 94 007 6973 1

Gnoli, Claudio (2008). "Ten long-term research questions in knowledge organization". *Knowledge organization*, v. 35, n. 2-3, pp. 137-149.
https://doi.org/10.5771/0943-7444-2008-2-3-137

Gnoli, Claudio (2016). "Classifying phenomena. Part 1: Dimensions". *Knowledge organization*, v. 43, n. 6, pp. 403-415.
https://doi.org/10.5771/0943-7444-2016-6-403

Gnoli, Claudio (2017a). "Classifying phenomena. Part 2: Types and levels". *Knowledge organization*, v. 44, n. 1, pp. 37-54.
https://doi.org/10.5771/0943-7444-2017-1-37

Gnoli, Claudio (2017b). "Classifying phenomena. Part 3: Facets". In: *Dimensions of knowledge: facets for knowledge organization*, Richard Smiraglia and Hur-Li Lee (eds.). Würzburg: Ergon, pp. 55-67. ISBN: 978 3 956502736

Gnoli, Claudio (2018). "Classifying phenomena. Part 4: Themes and rhemes". *Knowledge organization*, v. 45, n. 1, pp. 43-53.
https://doi.org/10.5771/0943-7444-2018-1-43

Gorman, Jamie C.; **Cooke, Nancy J.**; **Salas, Eduardo**; **Keyton, Joann**; **Beck, Stephenson J.** (2010). "Perspectives: Examining communication as macrocognition in STS". *Human factors*, v. 52, n. 2, pp. 335-339.
https://doi.org/10.1177/0018720810371338

Hjørland, Birger (2012). "Is classification necessary after Google?". *Journal of documentation*, v. 68, n. 3, pp. 299-317.
https://doi.org/10.1108/00220411211225557

Hjørland, Birger (2014). "Information science and its core concepts: Levels of disagreement". In: F. Ibekwe-SanJuan & T. M. Dousa (eds.). *Theories of information, communication and knowledge: A multidisciplinary approach* (pp. 205-235). New York: Springer. ISBN: 978 94 007 6973 1

Integrative Levels Classification. ISKO.
http://www.iskoi.org/ilc

Jacob, Elin K. (2004). "Classification and categorization: A difference that makes a difference". *Library trends*, v. 52, n. 3, pp. 515-540.
https://www.semanticscholar.org/paper/Classification-and-Categorization%3A-A-Difference-a-Jacob/544f3fbb77f9d2b414daa69e26de0960facc1438

Leon manifesto (2007).
http://www.iskoi.org/ilc/leon.php

Lyttle, Melanie A.; **Walsh, Shawn D.** (2018). "Leaving dewey for BISAC". *Public libraries online*, Nov. 6.
http://publiclibrariesonline.org/2018/11/leaving-dewey-for-bisac

Marty, Paul F. (2014). "Digital convergence and the information profession in cultural heritage organizations: Reconciling internal and external demands". *Library trends*, v. 62, n. 3, pp. 613-627.
https://doi.org/10.1353/lib.2014.0007

Olson, Hope A. (2007). "How we construct subjects: A feminist analysis". *Library trends*, v. 56, n. 2, pp. 509-541.
https://doi.org/10.1353/lib.2008.0007

Ranganathan, Shiyali R. (1962). *Elements of library classification.* Bombay: Asia Publishing House.

Repko, Allen; **Szostak, Rick** (2020). *Interdisciplinary research: Process and theory*, 4th ed. Thousand Oaks, CA: Sage Publications. ISSN: 978 1 412988773

Šauperl, Alenka (2009). "Precoordination or not?: A new view of the old question". *Journal of documentation*, v. 65, n. 5, pp. 817-833.
https://doi.org/10.1108/00220410910983128

Smiraglia, Richard P. (2001). *The nature of "a work": Implications for the organization of knowledge*. Lanham, MD: Scarecrow Press. ISBN: 978 0 810840379

Svenonius, Elaine (1993). "Pre-coordination or not? Subject indexing in the 90s: principles and practices". In: Holley, R. P.; McGarry, D.; Duncan, D.; Svenonius, E. (eds.). *IFLA Satellite Meeting*. Lisbon, 17-18 August. Ubcim Publications, v. 15. SAUR. ISBN: 3 598 11251 3

Svenonius, Elaine (2004). "The epistemological foundations of knowledge representations". *Library trends*, v. 52, n. 3, pp. 571-587.
https://www.ideals.illinois.edu/bitstream/handle/2142/1691/Svenonius571587.pdf

Szostak, Rick (2003). *A schema for unifying human science: Interdisciplinary perspectives on culture*, Selinsgrove PA: Susquehanna University Press. ISBN: 978 1 575910604

Szostak, Rick (2004). *Classifying science: Phenomena, data, theory, method, practice*, Dordrecht: Springer. ISBN: 978 1 4020 3095 6

Szostak, Rick (2011). "Complex concepts into basic concepts". *Journal of the American Society for Information Science & Technology*, v. 62, n. 11, pp. 2247-2265.
https://doi.org/10.1002/asi.21635

Szostak, Rick (2012). "Classifying relationships". *Knowledge organization*, v. 39, n. 3, pp. 165-178.
https://www.nomos-elibrary.de/10.5771/0943-7444-2012-3-165.pdf

Szostak, Rick (2017). "Facet analysis without facet indicators". In: Smiraglia, Richard; Lee, Hur-li (eds.). *Dimensions of knowledge: Facets for knowledge organization*. Wurzburg: Ergon, pp. 69-86. ISBN: 978 3 956502736

Szostak, Rick (2019). *Basic Concepts Classification. ISKO Encyclopedia of knowledge organization.*
https://www.isko.org/cyclo/bcc

Szostak, Rick; **Gnoli, Claudio**; **López-Huertas, María** (2016). *Interdisciplinary knowledge organization*. Berlin: Springer. ISBN: 978 3 319 30148 8

United Nations Standard Products and Services Code (*Unspsc*) (n.d.).
https://www.unspsc.org