Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

667

Notation†

Claudio Gnoli

University of Pavia – Science and Technology Library via Ferrata 1, Pavia 27100, Italy,
<claudio.gnoli@unipv.it>

Abstract: Notations are systems of symbols that can be combined according to syntactical rules to represent meanings in a specialized domain. In knowledge organization, they are systems of numerals, letters and punctuation marks associated to a concept that mechanically produce helpful sequences of them for arranging books on shelves, browsing subjects in directories and displaying items in catalogues. Most bibliographic classification systems, like *Dewey Decimal Classification*, use a positional notation allowing for expression of increasingly specific subjects by additional digits. However, some notations like that of Bliss *Bibliographic Classification* are purely ordinal and do not reflect the hierarchical degree of a subject. Notations can also be expressive of the syntactical structure of compound subjects (common auxiliaries, facets etc.) in various ways. In the digital media, notation can be recorded and managed in databases and exploited to provide appropriate search and display functionalities.

1.0 Introduction

Notations are systems of written symbols that can be combined according to some set of syntactical rules to represent various meanings in a specialized domain. Familiar examples include mathematical or logical formulas using numbers, variables and operators; formulas denoting chemical compounds by the kind, number and bonds of their atoms; and successions of notes forming a musical score. Such systems can be understood as special languages, that is languages for special purposes, and as artificial languages (Sammet and Tabory 1968). They are typically alternative to the expression of equivalent contents in words, which was more common in former literature; occasionally, words themselves may be used as in the "verbal notation" (2011) for music.

Several more specialized domains, including knowledge organization, have also developed their own notations. For example, the Pfafstetter Coding System allows for ordering of river basins and their branches by a decimal positional notation (Verdin and Verdin 1999); the International Phonetic Alphabet allows for precise representation of pho-

nemes and their sequences in any natural language; programming languages for computers use various symbols for instructions and variables; the Laban notation is used to represent successions of movements of the human body in physical activity or in dance; chess matches are recorded by an algebraic notation indicating pieces and coordinates in the gameboard; pacenotes are recorded by special symbols in a notebook then read to rally drivers in order to anticipate the coming bends, junctions and optimal gears.

Bawden (2017) considers "the extent to which information representation and communication [of molecular structures by notation] has gone hand-in-hand with the development of concepts and theories in chemistry, so that it is difficult to tell where the one ends, and the other begins." He echoes Grolier (1991, 99-100) where he observed that "historians of science repeatedly assert that progress in such sciences as logics, mathematics and chemistry was largely conditioned by important innovations in notation (symbolization). The same judgment could be valid for classification." Bibliographic classification systems are indeed another important domain where notation

668

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

is applied. This article discusses notation in classification systems and, more generally, knowledge organization systems (KOS).

1.1 Historical precedents

Although detailed notational systems for knowledge organization have been developing especially since the nineteenth century, various precedents may be found in the earlier history of culture, that must have been influential for at least the very idea of representing and organizing concepts according to numerical or literal symbols. Only a very short mention of some of them is given here.

The ancient Judaic tradition of Kabbalah already associated concepts to letters, words and numbers mentioned in the Bible. This may have influenced such folk traditions as the Southern Italian association between objects or persons dreamed and the numbers one to ninety to be taken out in the lotto gambling. In an oral tradition of Naples, the meanings of numbers taken out progressively in the game can in turn be combined by a gay man (*femminiello*) to

create and develop a story. Association between numbers and concepts is also reflected in nursery rhymes listing relevant phenomena (one is the Sun, two are the eyes...) that can be seen as classifications *ante litteram*.

Medieval systems of knowledge organization used some forms of notation for purposes of mnemotechnics and learning of wisdom (Rossi 2000). Ramon Llull's *Arbre de filosofia desiderat* (1290) described a tree of knowledge including nine "flower" categories represented by letters "B" to "K": e.g. "B" "goodness, difference, power," "C" "magnitude, concordance, object," "D" "duration, contrariness, memory," etc. By rotating a wheel where such categories are written (Figure 1), these can be combined with nine more "branch" categories represented by letters "L" to "U" to give such combinations as "DP" "memory: unity or plurality" and "DS" "memory: similarity or dissimilarity."

Such tools for artificial memory and representation were cultivated again by Giordano Bruno (1548-1600) then by Johann Heinrich Bisterfeld (1661), who developed a "philosophical alphabet" associated to tables of terms and concepts of all sciences, including general categories
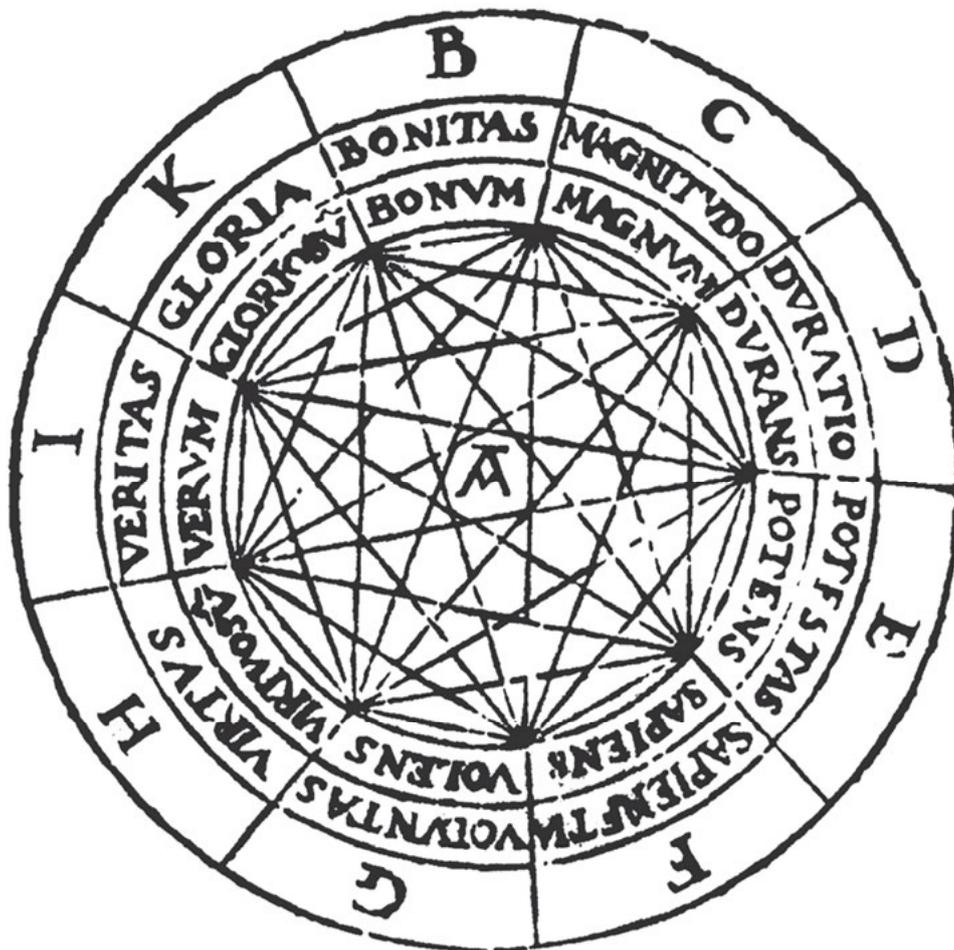

Figure 1. https://commons.wikimedia.org/wiki/File:Ramon_Llull_-_Ars_Magna_Fig_1.png.

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

669

and a "tabula primitiva" of what we would call today common auxiliaries. In Cave Beck's Universal Character (1657), the terms of language were listed and notated by digits "0" to "9" and combinations of them, to produce a "numeric dictionary" and an "alphabetical dictionary," each referring to the other much like in the relative index of a modern classification scheme. John Wilkins's Real Character (1668) famously used letters to identify elementary concepts listed in the hierarchical schedules of his "philosophical language," briefly described in a famous essay by J.L. Borges (1952) and discussed the next year by information scientist Brian Vickery (1953). A similar classification system with a literal notation that made it a true artificial language was George Dalgarno's Ars Signorum (1661):

Skam	grace
Skan	happiness
Skaf	to worship
Skab	to judge
Skad	to pray
...	

G.W. Leibniz's *Dissertatio de arte combinatoria* (1666), influenced by Bisterfeld (Loemeker 1961), suggested to associate numbers to elementary concepts ("1" point, "2" space, "3" between, "4" contiguous...) and to combine them into an algebra of all possible subjects, the Characteristica Universalis, although he did not develop it into a full system. Leibniz's ideas have been studied by logician and linguist Louis Couturat (1868-1914) who also developed Ido, an international auxiliary language.

1.2 Notation in modern knowledge organization systems

In the context of modern knowledge organization, notations are systems of symbols that identify the concepts of a KOS (Vickery 1952-1959; Daily 1956; 1976, 194 ff.; Grolier 1956; Coates 1957; 1959; Dobrowolski 1962; Mills 1967; A.C. Foskett 1996). Bliss (1940) described notation as "a system of symbols for maintaining the structural order of a classification and for locating terms, or subjects, in the classification," and Ranganathan (1945) as "an artificial language of ordinal numbers for the specific purpose of mechanizing arrangement." Ranganathan also makes a clear distinction between (1967, 327-8):

– the "idea plane," that is the concepts and relationships in a KOS,
– the "verbal plane," that is their expression in terms of some natural language, and
– the "notational plane," that is their translation into the symbols of some notation:

Along with the capacity to create ideas, came also the capacity to develop an articulate language as medium for communication. ... But, language is more lethargic than idea. Homonyms and Synonyms, therefore, grow like weeds. Undertones and overtones grow in abundance.

Therefore, attempts are continually in progress to make a language precise—at least among those creating ideas in a specific discipline. It is so at least for newly created ideas. Further, words are often replaced by symbols pregnant with precise meaning. When arrangement is found necessary, ordinal numbers are used as helpful symbols. A distinctive contribution of classification, as found and as being cultivated in the field of Library Science, is the Notational Plane. Uniqueness of the idea represented by an ordinal number and the total absence of homonyms and synonyms are the distinctive features of the notational plane, when compared to the verbal plane.

Notation is typical of classification schemes, while in such verbal KOSs as subject heading lists, thesauri, taxonomies and ontologies, concepts are primarily identified on the verbal plane through controlled "terms" formed with one or more words. However, notations can sometimes be used as well to represent concepts that are also identified by terms, for example as language-neutral identifiers in multilingual thesauri, or as record identifiers: e.g., in Medical Subject Headings (MeSH), the term "retina" can also be represented by its notation "A09.371.729," a subdivision of "A09.371" which represents the broader term "eye."

Homonymy and synonymy can also be managed on the verbal plane (unlike the quote above appears to suggest) in thesauri, but terms representing concepts do not include information on their ordinal and hierarchical position in the structure of the system. Indeed, in verbal systems, terms are usually presented in alphabetical order, which makes them easy to be searched only when the appropriate term is known in advance. On the other hand, as users do not always know an appropriate term by which their information need is expressed, a systematic arrangement according to some principle can also be useful to guide them across the collection of available documents. For some kinds of concepts, systematic arrangement is even required by common sense, as it would be inconvenient to list, e.g. "Friday," "Monday," "Saturday," "Sunday," "Thursday," "Tuesday," "Wednesday," or "divorce," "engagement," "marriage," "separation" in alphabetical order only.

In classification schemes, a systematic order is the preferred way of displaying concepts, while an alphabetical index (commonly known as the Relative Index in the *Dewey*

670

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

Decimal Classification, *DDC*) is only an auxiliary tool for finding the place of a concept in the systematic schedules. In order to control the systematic sorting of items indexed by a classification scheme, some notation is required (Ranganathan 1967, chapter HA). This feature may even be seen as the most typical to distinguish classification schemes from such other KOS types as taxonomies (where concepts also form hierarchical trees, but sister branches are listed alphabetically) or thesauri (where concepts are primarily listed alphabetically and hierarchical trees can only be inferred through series or BT/NT relationships). Unlike one may believe at first sight, the most important function of notation (see Section 3.0) is not to represent the corresponding concepts in a short form but to record the appropriate sequence in which they are presented, both in the schedules and in any set of information resources. This makes the notational system adopted in a KOS, with its peculiar properties, less trivial than the bare use of any set of abbreviations.

2.0 Representing notation

Within a document, numerical notations (see Section 3.0) such as those of the *DDC* or the Universal Decimal Classification (UDC) can usually be distinguished from bulk text as they consist (mostly) of numerals rather than letters. However, ambiguities may occur as numerals can also be used to represent quantities, document sections or other information. This is even more the case with notations that mainly use letters, such as that of Bliss *Bibliographic Classification* (*BC*).

To avoid ambiguity and express the nature of notation, then, this can be represented in a font different from bulk text. In some card catalogues subject-related headings were written in red, a heritage of "rubrication" (from Latin *rubrum* "red") of emphasized parts of old manuscripts. In modern digital-based printing and visualization on screen, no standard use has spread yet. Easy options are italics or bold as opposed to regular font. We recommend use of a monospaced font (such as Courier), as commonly adopted for representing code in computer science literature and for rendering the content of the <code> HTML element. This choice expresses the fact that notation is a special technical language other than natural language which forms the bulk of a text. An example of this use follows (Gnoli et al. 2011, 201):

> the facet `mqvtn2` "whales, in *area*" is seen as both a subclass of `mqvtn` "whales" and a subclass of 2 "in *place*" ... Also notice that the facet name, "*area*," has been recorded here as an alternative label. An alternative approach for facets could be the use of `skos:collection` classes

We adopt this use in the present article and the whole *ISKO Encyclopedia of Knowledge Organization*; the same style is adopted throughout ISKO 2010 proceedings (Gnoli and Mazzocchi 2010). In controlled vocabularies, the function of identifying concepts is played by controlled terms, so these can also be represented in a monospaced font. For the verbal captions that illustrate the meaning of a class notation, no standard use has spread either. To avoid ambiguity, these should also be distinguished from bulk text in some way. Vickery (1956) uses small capitals, brackets or quotation marks, as in the example above, are other easy options.

3.0 Notational bases

In principle, any set of written symbols may be adopted as a notation. A binary system, for example, may adopt only "0" and "1," or a red dot and a blue dot, or—and—like in the *I Ching* classic Chinese text. However, only letters and numerals have conventional orders that are widely known, which has obvious advantages for the ordering function often played by notation.

As many important modern classifications have developed in western culture, Roman letters or Hindu-Arabic numerals are the most common choices. Additional symbols like punctuation marks are sometimes added, especially since the development of UDC and Colon Classification (CC), although their standard sequence is less obvious and needs to be defined explicitly by developers then learned by users. In general, exceedingly complex notational bases are considered to be a hindrance to users, as parodied in the character of Sariette, a family librarian from a tale by Anatole France (1914) who devised so complex shelfmarks that they could only be understood by himself (Gnoli 2006).

3.1 Positional notation

DDC took its very name from the adoption of Hindu-Arabic numerals "0" to "9." They make it a "decimal" system not just in the sense that classes are subdivided into arrays of ten subclasses but especially in the sense that the resulting numbers must be read and interpreted in the same way as decimal numbers, according to the "positional notation" used in mathematics (as opposed to "sign-value notation" like in Roman numerals) extended to the "radix fractions" that can follow the decimal point. That is, despite 123 is greater than fourteen, 0.123 precedes 0.14, because two precedes four.

This practice opens the room for indefinite expansion of notation and of classification schedules themselves, as more characters specify more detailed subdivisions of a field of knowledge (Visintin 2005). Positionality allowing for indefi-

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

671

nite expansion of subjects can be considered to be a major technical innovation in the history of bibliographic classification (Gorman 1998, 106 emphasis original).

> Every day, in libraries throughout the world, cataloguers perform a feat of dazzling intellectual audacity. They classify books and other materials. In other words, they reduce the infinite dimensions of knowledge to a straight line from *000* to *999* or *A* to *Z*. There is an old cartoon of a gamekeeper and a fisherman. The first says "You can't fish here" to which the fisherman replies "I *am* fishing here." Classification, the thing that cannot be done, is done all the time by librarians. The amazing thing is that it works—classification numbers, those dots on the straight line, enable library users to locate materials and groups of materials with great ease and are used more and more in online systems to provide sophisticated subject access.

The positional principle is usually adopted already for notating the main classes of a scheme and their immediate subdivisions, although *DDC* requires that a class number has at least three digits, with the digit characteristic of every main class followed by "-00" (e.g., "300" rather than just "3" for "social sciences") and the two digits of their hundred subdivisions followed by "-0" (e.g., "380" rather than just "38" for "commerce"); however, this horror vacui is only a graphic convention with no effect on the system structure, and has indeed been successfully abolished in UDC. If more than three degrees of subdivision are expressed, the first three digits are followed by a dot, then by further digits in any number according to the subject specificity:

300	social sciences
380	commerce, communication, transportation
386	inland waterway & ferry transportation
386.4	canal transportation
386.40	[special subdivisions of canal transportation]
386.404	special subjects in canal transportation
386.4042	activities and services [in canal transportation]
386.40424	freight services [in canal transportation]

Such further digits used to be written in *DDC* by groups of three separated by a blank space for the sake of readability ("386.404 24"), but in the digital environment, blank spaces tend to be abandoned ("386.40424").

DDC was also an early application of the principle of relative shelving, as shelfmarks were now assigned to books themselves rather than to shelves (Figure 2). A book could now be assigned a shelfmark according to its subject and keep it regardless of its material position in shelves and rooms. This makes interpolation of shelfmarks possible, expressing more specific subjects, indefinite addition of new books by moving the adjacent books to the next shelf, or even move of a whole collection to a new place without changing its shelfmarks. Classmarks can also be detached from the shelving function, to denote the subject of a book in an abstract sense, be it used to define its position in a shelf or not, for example in a catalogue or a bibliography.

As mentioned above, the notational base of *DDC* was also adopted by UDC, which was originally created as a special version of *DDC*. UDC additionally introduced punctuation marks to specify common auxiliaries (already used some decades earlier in Andreas Schleiermacher's Bibliographisches System, see Stevenson 1978, not to consider Bisterfeld's system discussed in Section 1.1) such as places, time periods, languages, forms of the document, etc. Thus, a "pure" notation of digits evolved into a "mixed" notation of digits and punctuation marks. While pure notations use only one kind of symbols, mixed ones use several of them, e.g., both literals and numerals.

Apart from possible ambiguities in the filing order of punctuation marks, the notational base of *DDC* and UDC is regarded as optimal, because Hindu-Arabic numerals are more widely known across the world than Roman letters, which are exclusive of some alphabets. Indeed, numerals are also adopted by the *Korean Decimal Classification* (*KDC*, see Oh 2012) and the *Nippon Decimal Classification* (*NDC*), which are derived from *DDC*, and the *Library-Bibliographical Classification* (*LBC* or *BBK*) changed its Cyrillic letters to numerals for the sake of internationalization (Sukiasyan 2017, Section 2.5.6). UDC numeral notation is widely used as a common language in the libraries of many Eastern European countries, where national alphabetical subject headings would be less effective as the local languages are spoken by a relatively low number of users, making the development and maintenance of subject heading lists economically disadvantageous. A pure numeral notation representing a completely different ordering of knowledge is adopted in Dahlberg's Information Coding Classification (ICC) (Dahlberg 2008).

The main alternative to numerals are letters of the Roman alphabet, in the filing order *A* to *Z*. This base is adopted for the main classes of Charles Ammi Cutter's Expansive Classification (EC) and for the first two divisions of the Library of Congress *Classification* (LCC). In both these systems, letters are followed by numerals for further subdivisions, although LCC numerals occupy a fixed length of four digits rather than having a positional function:

672

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

Figure 2. Monographs in multiple languages arranged by *DDC* shelfmarks from classes *658* "management" and *659* "advertising and public relations" at the Humanities and Education Library, University of Udine, Italy (photo by Carlo Bianchini).

L	education (general)
LB	theory and practice of education
LB1705-2286	education and training of teachers and administrators
LB1771-1773	certification of teachers

Other important general classifications using letters for their first subdivisions are Bliss *BC*, where capital letters form the majority of classmarks with only some numerals used to indicate common auxiliaries; and Ranganathan's *CC*, where main classes are expressed by capital Roman letters and combined with small case letters, Greek letters, numerals and punctuation marks to produce very expressive but complex classmarks. The developing Integrative Levels Classification (ILC) uses lower-case letters for main classes and their subdivisions, capital letters for deictics and numerals for facet indicators (cfr. Section 4.3).

3.2 Number of sister classes in one array

While being a practical solution, the adoption of Hindu-Arabic numerals or the Roman alphabet also entails important effects on the structure of a classification scheme. Notations based on numerals or letters have different "capacity" (Mills 1967, 42), so that systems may have up to ten or twenty-six main classes, 100 or 676 subclasses etc. depending on their notational base. But after all, why should every concept be always subdivided into ten or twenty-six specifications like in a Procrustean bed? Clearly, *DDC* main disciplinary classes are ten rather than eight or fifteen as an effect of the notational base, rather than for any intrinsic property of knowledge fields. This problem of "not following the path of nature, but adapting plants to author's own prescribed method" was noticed already by botanist John Ray (1627-1705) while commenting on Robert Morison's classification of plants[1] (Ray 1848, cited in Rossi 2000, 252). Experts in library classification have stated it again (Kyle 1959, 19, reference omitted):

Sayers says the classificationist takes the whole field of knowledge and "first divides it into a number of broad convenient areas, which he calls his main classes." "Convenience" seemed at one time to mean "notational convenience," so that having decided to

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

673

use 9 digits the classificationist aimed at dividing knowledge into 9 or more or less equal and distinguishable fields; or, having decided to use the alphabet he sought for roughly 26 divisions.

and (Mills 1967, 38):

> Notation should reflect order, not determine it. Bliss has said that it is "correlative and subsidiary." The systematic sequence of topics is the essence of library classification. Notation is only the mechanism which maintains that sequence; it should be considered only after the problems of sequence have been decided.

While developing classification schemes, editors try to mediate between practical requirements of notation and intrinsic requirements of subjects' structure in various ways, for example, by not using all available symbols when a lesser number of subdivisions has to be expressed. Three technical devices have also been adopted in various systems to make a notational base better reflect knowledge structures: centesimal notation, telescopization and sectorizing digits.

"Centesimal notation" is the use of two numerals instead of a single one to identify sister classes belonging to one and the same array, so that the base is expanded from ten to one hundred (or, in principle, to one thousand etc., or to 26^2, 26^3 etc.); however, this hampers expressivity (Section 4.3) and makes notation longer. It is adopted in LBC:

5	health care, medical sciences
53.0/57.8	clinical medicine
...	
55.5	rheumatology
55.6	oncology
55.8	dermatovenerology
56.1	neuropathology, neurosurgery, psychiatry
56.6	stomatology
56.7	ophthalmology
56.8	otorhinolaryngology
56.9	urology
57.0	medical sexology
57.1	gynecology
57.3	pediatrics
...	

and occasionally in other systems for such arrays consisting of many subclasses as plant families in UDC:

582.7/8	Rosidae
...	
582.73	Fabales
582.74	Sapindales
582.75	Geraniales
582.77	Myrtales
582.79	Apiales
582.82	Vitales

"Telescopic notation" is the conscious squeezing of two degrees of subdivision into a single notational array, like in the following example (Bhattacharyya and Ranganathan 1978, 139):

I1	Cryptogamia
I2	Thallophyta
I3	Bryophyta
I4	Pteridophyta
I5	Phanerogamia
I6	Gymnosperm
I7	Monocotyledon
I8	Dicotyledon

or in *DDC*:

722-724	architecture schools and styles
722	architecture from origins to 300 AD
723	architecture 300 to 1499 AD
724	architecture after 1400
725-728	specific structure types
725	public structures
726	building for religious purposes
727	buildings for education and research
728	residential and related buildings

On the other side, a class may occasionally need to be divided into more than ten or twenty-six subclasses. A common case is a list of the twelve months in a year represented in a numeral notation: 1 "January," 2 "February," ..., 8 "August," ?? "September," ... To deal with such problems, Ranganathan defined the practice of "sectorizing digits" or "empty digits," consisting in keeping the first and/or last digit of a notational base for expansion of the preceding notation (Ranganathan 1967, 312-313, cross references omitted):

> Another method of satisfying the Canon [of extrapolation in array] is to postulate the first and the last digits to be empty—for use as sectorizing digits. This method will admit of extrapolation at the beginning and at the end respectively of the array.

> In *DCC* and UDC, the digit "0" and the digit "9"—the first and the last digits of the pure base of Indo-Arabic numerals—are used as sectorizing digits in some arrays ....

674

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

CC uses a mixed base. Each species of digits forms a zone in an array. There is a sectorizing digit for each zone—"z" for Roman smalls, "9" for Indo-Arabic numerals, and "Z" for Roman capitals. Thus, it provides for any number of extrapolations at the end of each zone of an array. Further, extrapolation at the end is also possible by using packet notation. This amounts to extrapolating a whole zone at the end ....

Example in zone (Z—1)

R6   Indian philosophy
R68   Dvaita philosophy (Dualism)
R691   Charvaka philosophy (Materialism)

While empty digits expand an array at its end, "emptying digits" allow for "interpolation of a new number between any two existing class numbers or isolate numbers" (Ranganathan 1967, 314). This can be observed even in CC main classes, where main class "KX" "animal husbandry" has been interpolated between "K" "zoology" and "L" "medicine." A similar device has been proposed by Farradane (1952), consisting in expanding notation by introducing a different kind of symbol in the same array, e.g. "J," "K," "K1," "K2," "L," "M"..., just like Latin terms *bis*, *ter* etc. are sometimes added to a number to interpolate further items in a list. Of course, this makes notation less consistent and elegant.

While the exact number of symbols available for subclasses depends on the historical accidents of writing systems, the general fact that subclasses are associated to one or few tens of symbols may have natural bases. Indeed, all humans tend to group items into sets of manageable size for cognitive and practical purposes. The very fact that we categorize phenomena by a finite number of words, grouping them into classes instead of using a different symbol for every individual phenomenon, is a basic cognitive function. Miller (1956) famously identified "the magical number seven, plus or minus two" as the average number of items that can be processed by working memory. Within library and information science, Blair (1980) introduced the important notion of "futility point," that is the maximum number of items that the average user is willing to browse before changing her search strategy. Wiberley, Daugherty and Danowski (1990) found that most users of library catalogues examine some thirty to thirty-five items when displaying lists of search results. Bates (1998) notices that this agrees with the Resnikoff-Dolby Rule, according to which the average ratio between a book title and the corresponding table of contents, that between the table of contents and the back-of-the-book index, that between this and the full text, and those between several other information units all are very close to 29.55 (Resnikoff and Dolby 1972).

The fact that these values approach thirty, a number not very different from that of letters in many alphabets, also suggests that alphabets themselves may have originally developed as tools helping to organize words and concepts into manageable groups. "The Resnikoff and Dolby research also clearly needs to be related to the research on menu hierarchies in the human computer interaction literature" (Bates 1998). Therefore, an alphabetical or numerical notational base is after all not a completely arbitrary device in epistemological terms. The matter looks different in ontological terms, as there seems to be no special reason why phenomena should occur in arrays of few tens—known chemical elements are about one hundred, nitrogenous bases forming DNA are just four, spoken languages are several thousands, etc.

3.3 Pronounceable notations

Being designed for the prior purpose of controlling class order, in most cases notation is unsuitable for direct pronunciation, even when letters are used as the resulting sequences often include many adjacent consonants. Still, a pronounceable notation can be useful for oral communication of subjects and easier memorization (Cordonnier 1944; 1951, 27-29; Grolier 1953; 1956; Vickery 1956, 78-79). Pronounceable notations have sometimes been suggested, already in the last years of the nineteenth century by Verner and by Ricci (Kervégant 1962, 75-76; Dobrowolski 1964, 140-143), reminiscent of the philosophical languages of the past. D.J. Foskett and J. Foskett (1974) designed a special faceted classification for education where consonants and vowels always alternate, so that notation can directly be pronounced:

L   teaching method
Lim   direct method

M-P   curriculum
Men   French

R-S   educands and schools
Rid   secondary modern school

Rid Men Lim: direct method, French, secondary modern school

However, this requires a notational base only consisting of letters and tends to produce long classmarks that are only reasonable in simple, domain-specific schemes. The alternative solution is to establish rules by which symbols, even including numerals and punctuation marks, can be pronounced. Dobrowolski (1964, 141) proposes that the same digit is pronounced differently according to its position, so

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

675

that sequences of consonants followed by vowels are always produced; clearly this requires users to learn the rules. Recently, ILC has adopted pronunciation rules as a secondary feature not influencing the sequence of digits in notation itself (Gnoli 2018).

4.0 Functions of notation

Notations usually perform several functions at once and at the same time. These are sometimes described as the "qualities" of a good notation (Kervégant 1962, 68-76). In this section, the different functions are discussed separately, starting with the most fundamental ones.

4.1. Concept identification

Using a notation means to identify a concept within a KOS in a precise, concise way, independently from the vocabulary of any natural language. Indeed, as in Ranganathan's quote above, "[u]niqueness of the idea represented by an ordinal number and the total absence of homonyms and synonyms are the distinctive features of the notational plane."

The notation for a concept usually is shorter, although more cryptic, than its formulation in words; this makes notation useful for representing the concept in contexts where a limited space is available, like the spine of a book. Brevity indeed is often a desired quality in notation. Vickery (1956; 1957) developed sophisticated calculations to estimate the average length of different types of notations, concluding that the briefest notations should be purely ordinal rather than expressive (Section 4.2), should be retroactive (Section 4.3) unless the numbers of concepts per facet exceeds a certain limit, and should have distinctive symbols for main classes. However, in the contemporary digital context, expressivity has become more important than brevity (Section 7).

In synthetic systems, notation can identify a concept that is combined with others, such as a common auxiliary (=14 always means "Greek language" in UDC classmarks) or an isolate used as a facet. Classifications often adopt "parallel divisions" where two different classes can be divided into the same subclasses, like with geography of countries and history of the same countries, or with zoology of mammal groups and palaeontology of the same groups. For example, in *DDC*:

562   fossil invertebrates
563   various marine and coastal fossil invertebrates
564   fossil Mollusca and Molluscoidea
565   fossil Arthropoda
...
592   (living) invertebrates

593   various marine and coastal (living) invertebrates
594   (living) Mollusca and Molluscoidea
595   (living) Arthropoda
...

4.2. Ordering

As explained in Section 1.2, a primary function of notation is to produce meaningful systematic orders of the concepts it represents. While positional notation (Section 3.1) does produce a meaningful order, the latter can also be obtained without the former (Coates 1956). Such a case has been described as a "purely ordinal" notation (Vickery) or "group" notation (Ranganathan). For example, in both the first (*BC1*) and the faceted second (*BC2*) edition of Bliss *Bibliographic Classification*, notation is devised in such a way that, while subclasses and their facets get sorted in a standard citation order from general to specific, not always are the modulation of degrees of subdivisions and the articulation of facets reflected in the notation structure; indeed, some classes have the same number of letters than their subclasses—an extensive use of telescopization—or even a greater number (D.J. Foskett and J. Foskett 1991):

JC	administration of educational institutions
JCC	. buildings & equipment & services
JCC E	. . planning & design, architecture
JCC P	. . maintenance, repair
JCC Q	. . . cleaning
JCC R	. . . decorating
JCC T	. . renewal, conversion
JCC V	. . security
JCC Y	. . site & buildings, campus
JCD	. . . buildings

The production of an optimal systematic order of documents, by applying rules of general-before-specific, decreasing concreteness of facets, their inverted citation order (Section 6), anteriorizing common isolates (Section 7) etc. plays a very important function in allowing users to browse a collection in effective ways. Either by going directly to library shelves or by first searching in a catalogue, a user will arrive to some point in the linear sequence of documents. If notation has produced a good systematic order, the adjacent documents will deal with the same subject or with subjects related to the first one, thus allowing the user to continue exploration and to discover unknown resources of interest through serendipity. Ranganathan conceptualized this in his APUPA model, where a central umbral (U) document is surrounded on both sides by partly-relevant penumbral (P) ones, with a gradual transition to irrelevant alien (A) ones (Satija 2017, section 4.3.11; Giusti 2018).

676

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

4.3. Expressivity

In "expressive" notations, unlike purely ordinal ones, the structure of the KOS is reflected largely in the structure of classmarks (Broughton 1999):

– the number of digits in a class symbol corresponds to the degree of specificity of the represented concept, as in *DDC* and systems derived from it;
– synthetic combination of concepts is reflected in synthesis of notation pieces, as in UDC and CC;
– the categories to which concepts belong have a stable notational representation (e.g., energy facets are always introduced by a colon in CC; deictic classes are always represented by capital letters in ILC).

While not being mandatory as the previous section illustrates, expressivity of the KOS structure often is a desirable quality in notation. It can work as a cognitive guide for users paying attention to notation:

It seems that many people, not librarians, automatically assume that a notation should be expressive. If told, say, that *SE* is English Law (as it is in the BC) they will assume that any divisions of English Law begin with *SE* and are surprised to find that English Commercial Law, say, is *SL*. (Mills 1967, 40)

As combined with concept identification, expressivity is especially useful in digital information retrieval, as the same concept is always identified and retrievable by the same notation whatever its position within synthesized combinations. Furthermore, right truncation will allow for simultaneous search of a class together with all its subclasses, as these share the same initial characters: searching *DDC* class "386" "inland waterway & ferry transportation" will also retrieve "386.4" "canal transportation"; this is not possible with non-expressive notations, as searching for BC2 class "JCCY" "site & buildings, campus" will not retrieve its subclass "JCD" "buildings" (cf. the examples above).

Expressivity of syntax can be obtained by using special digits to mark the articulation point of a compound: in *DDC* "599.094" "mammals of Europe," the "0" marks the point where a common subdivision is appended and the following "9" indicates that it is a historical-geographical subdivision; in UDC, the symbols ":" and "+" mean that the following notation is meant to be, respectively, in some relation or in coordination with the preceding one, as in "1+2" "philosophy and religion;" in CC, the specific punctuation marks work as "facet indicators," that is stand for

the fundamental category to which the subsequent facet belong (personality, matter, energy, space or time) and expressive symbols are also available for phase relationships (bias, comparison, etc.).

This kind of notation is a good illustration of Ranganathan's original idea of facet analysis as the combination of pieces by means of bolts and screws like in a Meccano toy; indeed, bolts and screws are expressed by punctuation marks or other symbols. As notation should always produce a "helpful sequence" of faceted classes, facet indicators themselves should be devised in such a way that they produce a meaningful order of facets, e.g., of "increasing concreteness" in CC (time < space < energy < matter < personality). To this purpose, within a faceted classmark, facets have to be cited in the inverted order of schedules ("principle of inversion"), that is "P," "M," "E," "S," "T" in CC (Satija 2017, section 4.3.7).

While the order of punctuation marks may look ambiguous, numerals or letters are more effective. The FAKTS draft faceted classification of humanities (Broughton and Slavic 2007) represents a classical faceted structure by an expressive notation, where numerals stand for subclasses and capital letters are facet indicators; letters are chosen in such a way to produce the standard inverted citation order thing, part, property, process, operation, patient, agent, place, time, theory, as required by facet analytical theory:

590	religion
590A	theory and philosophy of religion
590A4	God, gods
590E	persons and objects in religion
590E3	persons in religion
590E31	originator, founder
5904	Buddhism
5904A4	gods in Buddhism
5904A443	physical form, appearance
5904E31	founder of Buddhism, the Buddha
5904E31A443	Trikaya, doctrine of three bodies of the Buddha

Expressivity often comes at a price: synthesized classmarks can get very long and clumsy, even redundant, as in "617.231:615.2-021.473" "carcinoid hearth disease" in UDC draft revision of medicine (McIlwaine and Williamson 2008, 11). Some concepts of very common use may have a longish notation as an effect of their logical position in the systematic schedules. A public library may have more documents on *DDC* "599.6655" "horses" than on its superclass "599.66" "perissodactyla." To manage this problem, it has been suggested that classes of middle specificity should have a shorter rather than a longer classmark than their parent classes, just as in natural languages frequently occurring words tend to be shorter. This require-

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

677

ment has been achieved in the original notational systems devised by Zygmunt Dobrowolski where, unlike classical systems, notation for the most general classes is synthesized from the first and the last of their subclasses (*symboles jumelées,* "twin symbols"). In this way, the middle degrees of specificity are represented by the shortest classmarks, while both their parent classes and their son classes have longer classmarks. (Dobrowolski 1964, 145, translated from French):

> As the classmarks of degree 1 and 2 are seldom used, the twin symbols will not be hindering for the practical use of the schedule indexed in such a way.

This can be illustrated with classmarks from a tree diagram by Dobrowolski (1964, 144), where main classes "0/3" and "4/6" have subdivisions with such shorter symbols as "1," "10" etc.:

0/6	most general class
0/3	first subclass of degree 1
0/1	first subclass of degree 2
0	first subclass of degree 3 [subclasses omitted]
1	second subclass of degree 3
10	first subclass of degree 4
11	second subclass of degree 4
2/3	second subclass of degree 2
2	first subclass of degree 3 [subclasses omitted]
3	second subclass of degree 3
30	first subclass of degree 4
31	second subclass of degree 4
4/6	second subclass of degree 1
4/5	first subclass of degree 2
4	first subclass of degree 3
5	second subclass of degree 3
50	first subclass of degree 4
500	first subclass of degree 5
501	second subclass of degree 5
51	second subclass of degree 4
6	third subclass of degree 3

Another way of achieving brevity at the expense of expressivity is dropping the initial characters of a class when it is used in a canonical combination with another one. ILC special facets allow this, as those initial characters are recorded in the schedules rather than expressed in the combined classmark; for example, "g" "continuum bodies" may have their colour specified by facet "g96;" foci for this facet are taken from class "darll," which notation is implied in the compound subject "g96i" "green bodies," that only keeps the "i" from "darlli" "green:"

darll	visible light
darlli	green
g	continuum bodies
g96 [darll]	colour
g96i	green bodies

Finally, if one completely renounces expressivity, a purely ordinal notation can be adopted, which is known to be brief from Vickery's calculations (Section 4.1). In this case, articulation between different parts of a compound notation, e.g., facets, can still be expressed by "retroactive" notation, a technique introduced by Barbara Kyle (1958, 171) and other members of the Classification Research Group and adopted systematically in BC2. In retroactive notation, apart from the first character, which expresses the main class, all subsequent characters must have an increasing ordinal value, e.g., a concept can have notation "JBFH" (as "B" < "F" < "H") and another "JDLR" (as "D" < "L" < "R"). When these are combined, as in "JBFHDLR," the point of articulation can be identified as the only place (..."HD"...) where the ordinal value decreases instead of increasing ("H" > "D"). This produces an elegant notation consisting in a single series of symbols.

One disadvantage of retroactive notation is that only a fraction of a notational zone is available for allocating concepts, e.g., class "JDL" can only be expanded by appending letters "M" to "Z," as earlier letters would be confused with another concept in compounds. Vickery (1957, 77; 1958b, 73) seems to consider retroactive notation to be the best option, as opposed to the expressive classical "signposted" notation, and Mills has indeed adopted it in BC2; on the other hand, this choice is exactly what makes this KOS unsuitable for exploitation in digital applications due to the lack of expressivity of its notation.

4.4 Mnemonicity

Notation can be mnemonic when a symbol suggests its own meaning in the schedules. Clearly this could easily conflict with the previously mentioned functions, so it is only implemented where it does not. Indeed, mnemonicity is usually not regarded as a primary function of notation. An exception is the Basic Concepts Classification (Szostak 2012), where main classes are represented by the first letter of the corresponding captions, thus renouncing even ordinality and producing a non-systematic order of classes: "A" "art," "C" "culture," "E" "economy," "G" "genetic predisposition." BC2 makes a more moderate use of this, as notation for "chemistry" is "C" and that for "physical anthropology, human biology, health sciences" is "H," but most other main classes have classmarks unrelated to their captions; the same happens in subclasses.

678

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

This association between notation and the first letter of its caption is not possible in classifications using numeral notations; still some numerals can also be associated more or less constantly to some meanings, as in Ranganathan's "seminal mnemonics." This can be observed already in *DDC*, where, e.g., "4" often (but far from always) means "France," "French" etc. Perhaps the most pervasive use of mnemonics (and expressivity of categories) is that found in ICC, where digits after the first one always correspond to one of the ten categories of the "Systematifier" by which a field of knowledge is divided into subfields (Dahlberg 2008):

...*0*   general form concepts
...*1*   theories, principles
...*2*   objects, components
...*3*   activities, processes
...*4*   properties, or 1st kind of field speciality
...*5*   persons, or 2nd kind of field speciality
...*6*   institutions, or 3rd kind of field speciality
...*7*   technology & production
...*8*   application in other fields, determination
...*9*   distribution & synthesis

For example, "823" "information handling" is an activity (..."3") on the objects (..."2") of science and information ("8"...). A similar recursive use of categories can also be found in ILC facet indicators.

5.0 Allocation of notation to concepts

In developing a classification, once the conceptual structure and a notational base are defined, concepts have to be assigned to arrays of symbols ("apportionment" or "allocation") (Mills 1967, 40). The first decisions concern the ordering of main classes. This depends on the dimensions privileged by the system (e.g., disciplines or phenomena), the delimitation of the system domain (e.g., general or special) and the inspiring philosophies (e.g., based on epistemological criteria as in *DDC* and LCC, or on ontological criteria as in BC2 and ICC). Choices clearly have philosophical implications on the resulting ordering of knowledge items, both in the broad outline of classes and in their deeper subdivisions. For example, Ranganathan opted for an original order of main classes based on a bell-shaped sequence of "increasing concreteness" peaking at Δ "spiritual experience" then continuing downwards with "increasing artificiality" (Bianchini et al. 2017; Satija 2017, sections 4.3.2-4.3.3).

The distribution of symbols can also be performed mechanically by algorithms populating the available arrays, thus avoiding unconscious biases towards some symbols: "a Distribution Dictionary is a predefined structured code list for each character based on frequency of occurrence; its purpose is to make notation short and evenly distributed" (Liu 1990, 18).

As knowledge evolves, every KOS needs to be updated from time to time. The internal organization of a class may change depending on the evolution of a domain, of its structure and of its terminology, which are studied as "subject ontogeny" (e.g., Tennis 2017). Also, new classes may emerge and become more important (Richmond 1958); an outstanding example is the explosion of literature on computer science, that has persuaded *DDC* editors to allocate more classes ("004-006") for it, instead of just very specific subclasses as before.

While the choice of notation for new classes should depend on their logical relationships with the existing ones, in longstanding classifications it is also biased by limitations in available notation. Expansion of classes according to progress and change in knowledge should always be possible in a "hospitable" notation, either at the extremes of an array ("extrapolation") of within it ("interpolation"). This was indeed one way how Ranganathan wanted to improve *DDC* solutions (Tunkelang 2009, side 7). Pragmatic requirements and editors' experience often suggest some strategies, like always leaving the first symbol in an array unassigned ("A" if using letters, "0" if using numerals) as it can later turn out to be useful for special purposes including extrapolation of new classes: otherwise, it will never be possible to interpose new classes before the first one, thus forcing editors to allocate them further down, which may not reflect the desired knowledge structure. Various devices can also be available to adapt notation to structure, such as emptying digits described in Section 3.2.

Classes can also be regrouped if a field is subsumed under a more general one or to improve consistence. A famous example is UDC class "4," which is empty after language (which was "400" in the original *DDC* source) has been unified with "8" literature. UDC editors and advisors often discuss about possible new uses of "4," and very diverse hypotheses have been submitted. In the literature on *DDC*, sections of notation that once had a meaning but are now free are known as "phoenix schedules," as they have died but can arise from their own ashes again, like the mythological bird, when a new edition is published; however, editors usually wait many years before using a phoenix schedule again, to prevent confusion in old collections still classed with the original meaning.

Some classes may be allocated in a position different from their logical structure even on purpose. This may be the case of subclasses that are considered to be especially important because of their frequent use or their relevance to a local situation. In the Soviet *LBC*, as well as the *Chinese Library Classification* (*CLC*) derived from it, the first main class "A" is for Marxism-Leninism. In CC, some digits are

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

679

reserved in all arrays to express the favoured host class ("0"), the mother country ("2," usually India in most applications of this scheme, but potentially a different country) and the favoured country ("3," usually UK in Indian libraries, but could, e.g., be USA in a UK library) (Ranganathan 1967, 130-131; Gatto 2006). *ILC* generalizes this use by reserving all capital letters "A" to "T" for main classes of locally favoured concepts.

6.0 Syntax

As notation can be composed of several parts, syntactical issues emerge. They are especially interesting with expressive notations (Section 4.3), which can be exploited to manipulate concepts in automatic ways.

When combining two concepts, the second concept is usually interpreted as a specification of the first one. For example, philosophy of science can be expressed in UDC as either "5:1," literally meaning science in some relation with (e.g., treated in the perspective of) philosophy, or as "1:5," meaning philosophy in some relation with (e.g., dealing with) science. The choice has effects on the arrangement of concepts, as documents classed under the former combination will be filed together with other documents on science, but documents classed under the latter will be filed together with documents on philosophy. Therefore, choice should depend on which aspect is prior in the document itself, although local preferences may also affect it. UDC also provides a "::" symbol to specify that the order of the combination is relevant and cannot be inverted. Similar combinations are possible in CC: "Z(Q7)" "Islamic law," "Z&gQ7" "law influenced by Islamic religion." The first component of such combinations has been described in alphabetical subject indexing as the "base theme" of the document, while further connected components can be "particular themes" (Cheti 1996); Gnoli (2018) applies these notions to classification and its notation.

In faceted systems, notation for the facets should usually be expressed according to the "facet formula" of the class, following a standard citation order (Wali and Koul 1972). In CC, this famously is "PMEST," that is personality facets should be cited first, followed by matter facets, then energy facets, space facets and time facets. As mentioned, notation should be devised in such a way that, in the schedules, the inverted order will appear, that is classes specified by a time facet only should be listed before these with a space facet, then those with an energy facet etc. This is implemented in a clear way in the FATKS example of Section 4.3 above. Vickery (1958a, 10) proposes a general citation order of the structural elements of a synthetic classification, which should be reflected in the order produced by notation.

While revising UDC class "2" for religion, Vanda Broughton has identified a notational "'Genesis' problem." This deals with concepts that are subdivisions of a faceted concept, like the book of "Genesis" should be a subdivision of the Bible in Judaism and other religions. UDC notation allows to express the concept "Bible" as a faceted combination "26-23" "Judaism, sacred books." Now, notation for "'Genesis'" should be a subdivision of the whole faceted compound "26-23," rather than a subdivision of "-23" sacred books in general, so that just adding a digit after "-23" would be inappropriate; UDC editors agreed that the system, like other known classifications, lacks a notational symbol to express this, and that one could possibly be introduced to give, e.g., "26-23,11" "'Genesis.'"

Very complex combinations of concepts in synthetic notations may produce ambiguities in their interpretation, of the kind "a:(b:c)" vs. "(a:b):c." To deal with this, some systems may use punctuation marks to properly group components of synthesized subjects. Clearly these punctuation marks have to be different from all other components of the notation; for example, UDC uses square brackets for grouping, which are different from parentheses used for common auxiliaries of form, place and ethnic group. The effect of such punctuation marks on the proper sorting of classmarks has to be considered in practical applications.

7.0 Digital applications

While popular in traditional collections of printed books, notation tends to be less prominent in digital catalogues and digital libraries, as retrieval of terms from any position in a text string can replace browsing of concepts in a systematic order (Markey, Mitchell and Vizine-Goetz 2006). Classification systems and their notation are still researched, but involve fewer scholars as compared to such KOSs typical of the digital age as taxonomies, folksonomies and ontologies. Some authors even believe that the transition to the digital has made notation obsolete (Mai 2004):

> The principles for the construction of bibliographic thesauri and classification systems often advise that a notation is created to connect the different parts of the thesaurus or classification system. A notation is superfluous on the Web since the access mechanism and the documents are part of the same system.

However, notation can still be useful in digital media for controlling display of subjects and of corresponding document descriptions in a systematic order (Figure 3), which is a cognitively important function (Slavic 2006). Although

680

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

Figure 3. Display of book records sorted by BC2 shelfmarks (in the red oval) in the online catalogue of Fitzwilliam College Library, Cambridge, UK, http://library.fitz.cam.ac.uk/ (retrieved 2018-01-16).

concepts can be retrieved from any position of a subject heading, the choice of the theme cited first in a compound determines how the document will be grouped with others sharing the first cited theme when listed in browsable interfaces or in search results displayed in systematic order (Gnoli and Cheti 2013).

It is not clear to what extent the diminished use of notation is due to a lesser need for it, rather than to poor investments in interfaces exploiting its full meaning. Indeed, most online library catalogues do not yet provide adequate ways to browse documents by subject, to explore relationships between subjects, to sort results by subject, or to navigate from an identified document to others sharing the same subject with it (Bland 2008; Casson, Fabbrizzi and Slavic 2011). This situation contrasts with IFLA's acknowledgment that "explore," that is "[t]o discover resources using the relationships between them and thus place the resources in a context," is one of the five basic user tasks in the *Library Reference Model* (Žumer 2017, section 3).

For example, displaying notation alone without the corresponding captions, as it happens in many OPACs, is hardly useful to the majority of users. As what matters to users is the meaning of notation rather than its technical details, it is captions that should be displayed with more prominence. Notation can even be hidden and only work

in the background as a mechanical device controlling items' sorting. However, displaying notation will provide users with an additional hint of how the system works, like in transparent-case watches, which can be productive in the long term if not in quick searches. Only some catalogues have invested more in visualizing classification in their interfaces: a well-known example is the OPAC of the Polytechnic of Zurich (ETH), which displays UDC notation and captions in three different languages (Pika and Pika-Biolzi 2015), thus implementing the function of classification as a conceptual bridge in multilingual contexts mentioned in Section 3.1.

As notation is a sophisticated device, to be understood and fully exploited by its users, it needs to be modeled adequately in the architecture of databases and presented accordingly in interfaces. Slavic (2008) has discussed how it should be managed and maintained properly in relational databases, including provision of fields to record information about hierarchical and associative relationships between classes, mapping with classes of similar meaning in previous editions, editing dates, etc.

In synthetic classification systems, the captions of compound classes are usually not listed in the schedules but have to be obtained dynamically by interpreting notation for every occurrence of a compound subject. This requires that synthetic notation is parsed and divided into its com-

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

681

ponents, so that the appropriate caption for each of them can be obtained from the schedules. Several algorithms and procedures have been developed to this purpose in the past decades, especially for UDC (Buxton 1990); a recent example is by Piros (2017).

Bibliographic databases, in turn, should import data from KOS databases, including notation, captions and the scheme and edition to which they belong, and keep them in separate fields. The MARC formats (https://www.loc.gov/marc/classification/) indeed provide for such fields (Wajenberg 1983), although in many applications, these are not used. SKOS, the format conceived to represent KOSs in the semantic web, has been conceived primarily for thesauri and enumerative classifications used in the USA, such as LCC and *DDC*. Although SKOS does allow for recording notation in the "skos:notation" element, such syntactical components of analytico-synthetic systems as facets, foci, sources of foci etc. cannot be represented and exploited adequately with it (Gnoli et al. 2011). Some have suggested that the greater flexibility of the OWL format could be used instead (Zeng, Panzer and Salaba 2010), but no standard OWL application for synthetic notation has been developed yet.

As for visualization, most computer systems sort strings by default according to character encoding based on the American Standard Code for Information Interchange (ASCII), also reproduced in ANSI character set, Microsoft's WGL4 character set and the Unicode international standard. This means that blank space (ASCII hexadecimal code 20) will precede every digit, then some punctuation marks will precede numerals "0" to "9," which in turn will be followed by other punctuation marks, capital letters "A" to "Z" and lower-case letters "a" to "z." In case a classification system prescribes a different order, this needs to be produced by a special script that is capable of processing notation strings. For example, a class span such as UDC "1/2" "philosophy to religion" has a more general meaning than its first term "1" "philosophy;" therefore, it should be displayed before it according to the basic principle that general concepts precede specific concepts. However, in ASCII order "1" will precede "1/" thus producing the reverse order. A similar problem occurs in CC with "anteriorizing common isolates," by which bibliographies, synopses, histories or glossaries on a subject, like "bibliographies of Indian literature," should be listed before the simple subject itself, like "Indian literature," as they are supposed to have an introductory function for users (Satija 2017, section 4.3.11); again, appended notation expressing the anteriorizing common isolate will be filed in automatic applications after the simple subject, despite what was intended by the classification designer.

Script languages able to process strings are widely available (e.g., JSP, PHP), and may be used to create algorithms that produce the correct sequences. Alternatively, the ordi-

nal value of simple and combined classes can be specified in a special field of the database and processed. However, these solutions introduce the additional requirements that: 1) a programmer develops a suitable code or compile a database field; and, 2) the code or field interpreter is incorporated in the information system. Such resources may not be available in such contexts as small indexing projects or sharing of metadata across multiple institutions and platforms. To prevent these kinds of problems, notation for new classification systems can be developed considering the ASCII value of digits used in notation since its origin. For example, ILC adopts capital letters to represent favoured classes to be filed before standard classes represented by lower-case letters because of their position in ASCII.

8.0 Conclusion

Notation is a fundamental component of classification systems, and sometimes an auxiliary component of other KOS types. Its main function is mechanical control of concept ordering. Such ordering has important cognitive consequences for users, even in the digital media. Additionally, expressive notations can allow control, both mechanical and digital (querying, extraction, sorting), of individual structural components of a classification system.

To these purposes, a notation should be devised in such a way to respect the principle of general before specific, both within structural components and between them. In general, class spans and anteriorizing common isolates should precede simple classes; phase relationships should precede facets, and these in turn should precede subclasses; locally favoured classes should precede standard classes.

In modern digital applications, notation requires special database fields, scripts and interface design, in order to produce the optimal sequence of its structural elements for effective browsing of knowledge items arranged in systematic order.

Note

1. Ray (1848, 41-42): "Dr. Morison in opusculo nuper edito, cui *Praeludia Botanica* titulum fecit, illas, illarumque tacito nomine autorem, an pro meritis an indignis modis excepit, aliorum judicium esto. Nec tamen mirum tabulas confusas erroneas et imperfectas esse, cum trium tantum hebdomadum opus fuerint, ego vero nihil antea ejusmodi destinaveram, nec de eo unquam cogitaveram. Praeterea in iis ordinandis coactus sum non naturae ductum sequi, sed ad autoris methodum praescriptam plantas accommodare, quae exegit ut herbas in tres turmas seu tria genera quamproxime aequalia

682

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

distribuerem, singulas deinde turmas in novem diffe-
rentias illi dictas h. e. genera subalterna dividerem, ita
tamen ut singulis differentiis subordinatae plantae cer-
tum numerum non excederent: tandem ut plantas una
binas copularem seu in paria disponerem. Quae jam
spes est methodum hanc absolutam fore et non potius
imperfectissimam et absurdam?"

References

Bates, Marcia. 1998. "Indexing and Access for Digital Li-
braries and the Internet: Human, Database, and Domain
Factors." *Journal of the American Society for Information Sci-
ence and Technology* 49: 1185–205.

Bawden, David. 2017. "Chemistry and its (Information)
History." *The Occasional Informationist: Irregular Thoughts on
the Information Sciences* (blog), September 3. https://theoc
casionalinformationist.com/2017/09/03/chemistry-
and-its-information-history/

Bhattacharyya, G. and Ranganathan, S.R. 1978. "From
Knowledge Classification to Library Classification." In
*Conceptual Basis of the Classification of Knowledge: Proceedings
of the Ottawa Conference on the Conceptual Basis of the Clas-
sification of Knowledge, October 1st to 5th 1971*, ed. Jerzy A.
Wojciechowski. 2nd. ed. New York: Saur, 119-43.

Bianchini, Carlo, Luca Giusti and Claudio Gnoli. 2017.
"The APUPA bell curve: Ranganathan's Visual Pattern
for Knowledge Organization." *Les cahiers du numérique*
13, no. 1: 49-68.

Bisterfeld, Johann Heinrich. 1661. *Alphabeti philosophici libri
tres*. Vol. 1 of *Bisterfeldius redivivus*. Hagas Comitus: Vlacq.

Blair, D. C. 1980. "Searching Biases in Large Interactive
Document-Retrieval Systems." *Journal of the American
Society for Information Science* 31: 271-77.

Bland, Robert N. and Stoffan, Mark A. 2008. "Returning
Classification to the Catalog." *Information Technology and
Libraries* 27: 55-60.

Bliss, Henry Evelyn. 1940. *Classes A-G: Philosophy, Mathe-
matics, and the Natural Sciences*. Vol. 1 of *A Bibliographic
Classification: Extended by Systematic Auxiliary Schedules for
Composite Specification and Notation*. New York: Wilson.

Borges, Jorge Luis. 1952. "El idioma analitico de John Wil-
kins." In *Otras inquisiciones (1937-1952)*. Buenos Aires: Sur.

Broughton, Vanda. 1999. "Notational Expressivity: The
Case for and Against the Representation of Internal
Subject Structure in Notational Coding." *Knowledge Or-
ganization* 26: 140-8.

Broughton, Vanda and Aida Slavic. 2007. "Building a Fac-
eted Classification for the Humanities: Principles and
Procedures." *Journal of Documentation* 63, no. 5: 727-54.

Buxton, Andrew B. (1990). "Computer Searching of UDC
Numbers." *Journal of Documentation* 46: 193-217.

Cheti, Alberto. 1996. "Manuale ipertestuale di analisi con-
cettuale," ed. Serena Spinelli.

Coates, Eric J. 1956. "Ordinal and Hierarchical Notation."
Classification Research Group Bulletin 1: 15-9.

Coates, Eric J. 1957. "Notation in Classification." *Classifi-
cation Research Group Bulletin* 1, no. 2: D1-D19.

Cordonnier, Gérard. 1944. "Classification et classement."
Bulletin d'information scientifique et technique 6: 41.

Cordonnier, Gérard. 1951. "Classification, classement, ra-
gement, et sélection." *CNOF: Revue mensuel de l'organisa-
tion* 25, no. 5-6: 19-31.

Dahlberg, Ingetraut. 2008. "The Information Coding
Classification (ICC): A Modern, Theory-Based Fully-
Faceted, Universal System of Knowledge Fields." *Axi-
omathes* 18: 161-76.

Daily, Jay E. 1956. "A Notation for Subject Retrieval Files."
American Documentation 7: 210-4.

Daily, Jay E. 1976. "Natural Classification." In *Encyclopedia
of Library and Information Science,* ed. Allen Kent, Harold
Lancour and Jay E. Daily. New York: Dekker, 19: 186-
206.

Dobrowolski, Zygmunt. 1964. *Étude sur la construction des
systèmes de classification*. Paris: Gauthier-Villars.

Farradane, Jason E. L. 1952. "Scientific Theory of Classi-
fication." *Journal of Documentation* 8: 73-92.

Foskett, Anthony C. 1996. "Notation." In *The Subject Ap-
proach to Information*. 5th ed. London: Library Associa-
tion, 183-99.

Foskett, Douglas J. and Joy Foskett. 1974. *The London Edu-
cation Classification: A Thesaurus-Classification of British Ed-
ucational Terms*, ed. D.J. Foskett and Joy Foskett. 2nd ed.
Supplement—Education Libraries Bulletin 6. London:
University of London Institute of Education Library.

Foskett, Douglas J. and Joy Foskett. 1990. *Bliss Bibliographic
Classification Class J: Education,* ed. D.J. Foskett and Joy
Foskett. 2nd ed. London: Bowker-Saur.

France, Anatole. 1914. *La révolte des anges*. Paris: Calmann-
Lévy.

Gatto, Eugenio. 2006. "Variazione locale e comunicabilità
globale." In *Classificare la documentazione locale: Giornata di
studio; San Giorgio di Nogaro (UD), 17 dicembre 2005,* eds.
Lorena Zuccolo and Claudio Gnoli. ISKO Italia. http://
www.iskoi.org/doc/locale4.htm

Giusti, Luca. 2018. "The Penumbra-Line: Ranganathan's
Journeys and the Genesis of APUPA Pattern." In *Chal-
lenges and Opportunities for Knowledge Organization in the Dig-
ital Age: Proceedings of the Fifteenth International ISKO Con-
ference, 9-11 July 2018, Porto*, ed. Fernanda Ribeiro and
Maria Elisa Cerveira. Advances in Knowledge Organi-
zation 16. Würzburg: Ergon, 380-91.

Gnoli, Claudio. 2006. "L'alfabeto e la sindrome di sariette."
AIDA Informazioni 24, no. 3-4: 111-6.

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

683

Gnoli, Claudio. 2018. "Classifying Phenomena. Part 4: Themes and Rhemes." *Knowledge Organization* 45: 43-53.

Gnoli, Claudio and Fulvio Mazzocchi, eds. 2010. *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the Eleventh International ISKO Conference 23-26 February 2010 Rome, Italy.* Advances in Knowledge Organization 12. Würzburg: Ergon.

Gnoli, Claudio, Philippe Cousson, Tom Pullman, Gabriele Merli, and Rick Szostak. 2011. "Representing the Structural Elements of a Freely Faceted Classification." In *Classification & Ontology: Formal Approaches and Access to Knowledge: Proceedings of the International UDC Seminar, 19-20 September 2011, The Hague, Netherland*, eds. Aida Slavic and Edgardo Civallero. Würzburg: Ergon, 193-205.

Gorman, Michael. 1998. "The Impossibilty of Classification." In *Our Singular Strenghts: Meditations for Librarians.* Chicago: American Library Association.

Grolier, Éric de. 1953. "l'etude des problèmes de classification documentaire sur le plan international." *Revue de la documentation* 20: 105-16.

Grolier, Éric de. 1956. "Nouvelles recherches sur la symbolisation." *Revue de la documentation* 23: 13-21.

Grolier, Éric de. 1991. "Some Notes on the Question of so-called 'Unified Classification'." In *Tools for Knowledge Organization and the Human Interface: Proceedings 1st International ISKO Conference, Darmstadt, Germany, August 14-17, 1990*, ed. Robert Fugmann. Advances in Knowledge Organization 2. Frankfurt M.: Indeks, 2: 85-108.

Kervégant, Désiré. 1962. *Introduction à la documentation agronomique: La classification bibliographique.* Annales de l'Institut National de la Recherche Agronomique. Paris: Institut National de la Recherche Agronomique.

Kyle, Barbara. 1958. "Towards a Classification for Social Science Literature." *Journal of the Association for Information Science and Technology* 9: 168-83.

Kyle, Barbara. 1959. "Examination of Some Problems involved in Drafting General Classification and Some Proposals for their Solution." *Revue de la documentation* 26: 17-21.

Liu, Songqiao. 1990. "Online Classification Notation: Proposal for a Flexible Faceted Notation System (FFNS)." *International Classification* 17: 14-20.

Loemeker, Leroy E. 1961. "Leibniz and the Herborn Encyclopedists." *Journal of the History of Ideas* 22: 323-38.

Mai, Jens-Erik. 2004. "Classification of the Web: Challenges and Inquiries." *Knowledge Organization* 31: 92-7.

Markey, Karen, Joan Mitchell, and Diane Vizine-Goetz. 2006. "Forty Years of Classification Online: Final Chapter or Future Unlimited?" *Cataloging and Classification Quarterly* 42, no. 3-4: 1-63.

McIlwaine, Ia C. and Williamson, Nancy J. "Medicine and the UDC: The Process of Restructuring Class 61." *Extensions and Corrections to the UDC* 30: 9-16.

Miller, G. A. 1956. "The Magical Number Seven, plus or minus Two: Some Limits on our Capacity for Processing Information." *Psychological Review* 63, no. 2: 81-97.

Mills, Jack. 1967. "Notation." Chap. 5 in *A Modern Outline of Library Classification.* London: Chapman & Hall.

Oh, Dong-Geun. 2012. "Developing and Maintaining a National Classification System: Experience from Korean Decimal Classification." *Knowledge Organization* 39: 72-82.

Pika, Jiří and Pika-Biolzi, Milena. 2015. "Multilingual Subject Access and Classification-based Browsing through Authority Control: The Experience of ETH-Bibliothek, Zürich." In *Classification & Authority Control: Expanding Resource Discovery: Proceedings of the International UDC Seminar 2015, 29-30 October 2015, Lisbon, Portugal*, eds. Aida Slavic and Maria Inês Cordeiro. Würzburg: Ergon, 99-110.

Piros, Attila. 2017. "The Thought behind the Symbol: About the Automatic Interpretation and Representation of UDC Numbers." In *Faceted Classification Today: Theory, Technology and End Users: Proceedings of the International UDC Seminar 2017, London (UK), 14-15 September*, eds. Aida Slavic and Claudio Gnoli. Würzburg: Ergon, 203-18.

Ranganathan, S. R. 1945. *Elements of Library Classification.* Kaikhushru Taraporevala Memorial Series in Library Science 4. Poona: N.K. Publishing House.

Ranganathan, S. R. 1967. *Prolegomena to Library Classification.* 3rd ed. Bombay: Asia.

Ray, John. 1848. *The Correspondence of John Ray*, ed. E. Lankester. London: Ray Society.

Resnikoff, H. L. and Dolby, J. L. 1972. *Access: A Study of Information Storage and Retrieval with Emphasis on Library Information Systems.* Washington, DC: U.S. Department of Health Education and Welfare, Office of Education, Bureau of Research.

Richmond, Phillis A. 1958. "A Simple Mnemonic Largebase Number System for Notational Purposes." *Journal of Documentation* 14: 208-11.

Rossi, Paolo. 2000. *Clavis universalis: Arti della memoria e logica combinatoria da Lullo a Leibniz.* 3rd ed. Saggi 244. Bologna: il Mulino. Reprint of the 2nd ed., 1983.

Sammet, Jean E. and Robert Tabory. 1968. "Artificial Languages." In *Encyclopedia of Library and Information Science,* eds. Allen Kent and Harold Lancour. New York: Dekker, 1: 632-57.

Satija, Mohinder P. 2017. "Colon Classification (CC)." *Knowledge Organization* 44: 291-307.

Slavic, Aida. 2006. "Interface to Classification: some Objectives and Options." *Extensions and Corrections to the UDC* 28: 24-45.

Slavic, Aida. 2008. "Faceted Classification: Management and Use." *Axiomathes* 18: 257-71.

684

Knowl. Org. 45(2018)No.8
C. Gnoli. Notation

Stevenson, Gordon. 1978. *Andreas Schleiermacher's Bibliographic Classification and its Relationship to the Dewey Decimal and Library of Congress Classifications.* Occasional Papers. Urbana: University of Illinois Graduate School of Library Science.

Sukiasyan, Eduard. 2017. "Library-Bibliographical Classification (LBC)." In *ISKO Encyclopedia of Knowledge Organization*, ed. Birger Hjørland, coed. Claudio Gnoli. http://www.isko.org/cyclo/lbc. I cannot find in *Knowledge Organization*.

Szostak, Rick. 2012. "The Basic Concepts Classification." In *Categories, Contexts and Relations in Knowledge Organization: Proceedings of the Twelfth International ISKO Conference, 6-9 August 2012, Mysore, India*, ed. A. Neelameghan and K.S. Raghavan. Advances in Knowledge Organization 13. Würzburg: Ergon, 24-30.

Tennis, Joseph T. 2017. "Facets and Change: Design Requirements for Analytico-Synthetic Schemes in Light of Subject Ontogeny Research." In *Faceted Classification Today: Theory, Technology and End Users: Proceedings of the International UDC Seminar 2017, London (UK), 14-15 September*, eds. Aida Slavic and Claudio Gnoli. Würzburg: Ergon, 163-68.

Tunkelang, Daniel. 2009. *Faceted Search.* San Rafael, CA: Morgan & Claypool.

"Verbal Notation." 2011. In *Word Event: A Verbal Notation Symposium, Bath Spa University, 21 September 2011.* https://wordevent.wordpress.com/home/verbal-notation/

Verdin, K. L. and Verdin, J. P. 1999. "A Topological System for Delineation and Codification of the Earth's River Basins." *Journal of Hydrology* 218: 1-12.

Vickery, Brian C. 1952. "Notational Symbols in Classification." *Journal of Documentation* 8: 14-32.

Vickery, Brian C. 1953. "The Significance of John Wilkins in the History of Bibliographical Classification." *Libri* 2: 326-43.

Vickery, Brian C. 1956. "Notational Symbols in Classification. Part 2: Notation as an Ordering Device." *Journal of Documentation* 12: 73-87.

Vickery, Brian C. 1957. "Notational Symbols in Classification. Part 3: Further Comparisons of Brevity." *Journal of Documentation* 13: 72-7.

Vickery, Brian C. 1958a. "Notational Symbols in Classification. Part 4: Ordinal Values of Symbols." *Journal of Documentation* 14: 1-11.

Vickery, Brian C. 1958b. "Notation for the Classified Catalogue." In *Classification and Indexing in Science.* London: Butterworths.

Vickery, Brian C. 1959. "Notational Symbols in Classification. Part 5: Signposted and Retroactive Notation, and Part 6: Pronounceable Retroactive Ordinal Notation." *Journal of Documentation* 15: 12-6.

Visintin, Giulia. 2005. "Punti su una Retta, Porte sull'Iperspazio." Blog. AIB-WEB. Contributi. http://www.aib.'it/aib/contr/visintin5.htm

Wajenberg, Arnold S. 1983. "MARC Coding of DDC for Subject Retrieval." *Information Technology & Libraries* 2: 246-51.

Wali, M. L. and R. K. Koul. 1972. "Development of Notation in Freely Faceted Classification: a Case Study." *Herald of Library Science* 11: 30-43.

Wiberley, S. E., R. A. Daugherty and J. A. Danowski. 1990. "User Persistence in Scanning Postings of a Computer-Driven Information System: LCS." *Library and Information Science Research* 12: 341-53.

Zeng, Marcia Lei, Michael Panzer and Athena Salaba. 2010. "Expressing Classification Schemes with OWL 2: Exploring Issues and Opportunities based on Experiments using OWL 2 for Three Classification Schemes." In *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the Eleventh ISKO Conference, Rome, Italy, February 23rd-26th, 2010*, eds. Claudio Gnoli and Fulvio Mazzocchi. Würzburg: Ergon, 356-62.

Žumer, Maja. 2018. "IFLA Library Reference Model (LRM): Harmonisation of the FRBR Family." *Knowledge Organization* 45: 310-8.