Alphabetization† ††

Wendy Korwin*, Haakon Lund**

*119 W. Dunedin Rd., Columbus, OH 43214, USA, <wendykorwin@gmail.com>
**University of Copenhagen, Department of Information Studies,
DK-2300 Copenhagen S Denmark, <hl@hum.ku.dk>

Wendy Korwin received her PhD in American studies from the College of William and Mary in 2017 with a dissertation entitled *Material Literacy: Alphabets, Bodies, and Consumer Culture*. She has worked as both a librarian and an archivist, and is currently based in Columbus, Ohio, United States.

Haakon Lund is Associate Professor at the University of Copenhagen, Department of Information Studies in Denmark. He is educated as a librarian (MLSc) from the Royal School of Library and Information Science, and his research includes research data management, system usability and users, and gaze interaction. He has presented his research at international conferences and published several journal articles.

Korwin, Wendy and Haakon Lund. 2019. "Alphabetization." *Knowledge Organization* 46(3): 209-222. 62 references. DOI:10.5771/0943-7444-2019-3-209.

Abstract: The article provides definitions of alphabetization and related concepts and traces its historical development and challenges, covering analog as well as digital media. It introduces basic principles as well as standards, norms, and guidelines. The function of alphabetization is considered and related to alternatives such as systematic arrangement or classification.

Received: 18 February 2019; Revised: 15 March 2019; Accepted: 21 March 2019

Keywords: order, orders, lettering, alphabetization, arrangement

† Derived from the article of similar title in the ISKO Encyclopedia of Knowledge Organization Version 1.0; published 2019-01-10. Article category: KOS general issues.

†† This article could not have been written without the engagement and help provided by the editor, Birger Hjørland. The authors also thank two anonymous peer-reviewers for valuable feedback.

1.0 Definitions and explanation

Alphabetization[1] is a kind of ordering. The *Oxford English Dictionary* (Oxford University Press 2018) defines ordering: 1a: "a. To place in order, give order to; to arrange in a particular order; to arrange methodically or suitably." Ordering may be understood in two ways:

1. arranging items in a sequence according to some criterion;[2]
2. categorizing: grouping items with similar properties.

It is the first of these meanings that is relevant in relation to the term "alphabetization." Besides alphabetical order as ordering criterion, other criteria such as chronological or systematic may be used for arranging items in a sequence.[3] Both these meanings of ordering are often used as synonymous with "sorting,"[4] although sorting is often preferred for mechanical procedures, such as sorting algorithms.[5]

In general, the most common uses of ordered (or sorted) sequences are:

- making lookup or search efficient;
- making merging of sequences efficient.
- enable processing of data (http://www.isko.org/cyclo/data) in a defined order.

Alphabetization is the process of establishing the alphabetical order of a set of items based on their names or headings.[6] Alphabetical order is the arrangement of items by sorting strings of characters[7] according to their position in a given alphabet.

In addition to conventions for ordering letters, other characters such as numbers, symbols, ideograms, logograms, and typographical issues such as lowercase and uppercase letter should be differentiated. The overall term for this is "alphanumeric arrangement."

Examples:

- Books can be arranged according to titles, authors, languages, and other characteristics displayed in headings (as can representations of books in catalogs).
- Back-of-the book indexes may contain alphabetically arranged names and keywords referring to the pages on which these names are mentioned or concepts corresponding to the keywords bring information.
- Entries in dictionaries and encyclopedias can be arranged according to headwords (in addition, indexes can be arranged according to derived or assigned terms or names).
- In reference lists (e.g., in this article and in all articles in the IEKO Encyclopedia) references are ordered alphabetically according to author and publication year.
- Computer sorted outputs from databases can be arranged according to many characteristics, including those mentioned in the examples above.
- Persons can be arranged according to their last names, first names, and occupations in a directory.
- Wine bottles may be arranged alphabetically in supermarkets according to, for example, country of origin or name of producer.

The process of alphabetizing headings starts by collocating those starting with the first letter in a given alphabet. Headings starting with the second letter are then collocated, and the process repeats through the last letter in the alphabet (in English this is mostly termed the A-Z order). Each collocated group is then arranged according to the second letter in the heading and so on, until the whole string of characters in the heading has been arranged (i.e., exact alphabetical order, cf. below).

Alphabetical order has been described as "unnatural and arbitrary" (Weinberger 2007, 26) rather than organic or intuitive. The reasons for this are:

1. alphabetical ordering is the ordering of items (or their representations) by the symbols used for their names or attributes. Because things and their attributes may have different names, a first kind of arbitrarity is involved;
2. because formal (rather than substantial) aspects are used in alphabetization, a second kind of "unnatural and arbitrary" order is involved. Books with similar titles might be kept together, even when they differ widely in their subjects. A translated book might also be separated from the same title published in its original language (although cataloging rules may apply the principle of uniform titles[8]).

Another issue arises with synonyms, which allow for the same concept to be expressed using different words (and therefore placed in different alphabetical locations). This is dealt with in library and information science by forms of "vocabulary control" (such as subject headings and thesauri). This issue will be dealt with in other articles in this encyclopedia.

Despite this "unnatural and arbitrary" order, alphabetization has proven itself extremely useful. It is a widespread practice valued for its ability to render large amounts of information readily accessible to users. Alphabetizing is such a firmly ingrained process in many cultures that users may scarcely notice the organizational scheme that helps them browse through record stores or locate icons on their computer desktop.[9] Its history reveals, however, that alphabetization was not an inevitable development, nor was it a practice adopted wholesale from the moment of its invention. Instead, it has existed alongside, and has frequently been combined with or challenged by, other means of arrangement.

As we shall see below, alphabetization is often a complex operation that demands much more knowledge than just the twenty-six letters in the English alphabet and their conventional order.

2.0 History

The literature on alphabetization is limited but related to large literatures on the developments of alphabets (e.g., Drucker 1995), writing systems (e.g., Diringer 1962; Hooker 1990; and Daniels and Bright 1996) and, at the broadest level, human symbolic evolution (Lock and Peters 1999). Each specific writing system has its own literature and may pose specific problems to the development of standards for its representation and ordering.

Alphabetization requires that letters bear consistent names and, most importantly, a standard order. In non-phonetic languages like Japanese and Chinese, for instance, alphabetization is less entrenched, as their logographic and syllabic characters support multiple arrangement possibilities. The English term "alphabet," on the other hand, embodies the very idea that it labels: the consistency and predictability of A, B, C. Michael Rosen explains that the word itself is constructed from the first two letters of the Greek alphabet, alpha and beta. He writes (2015, 395-6):

> The alphabet is then the "alphabeta," rather as if we were to call the number system the "one-two." Tracing the route back we go to Latin "alphabetum," back to the ancient Greek "alphabetos," back to Phoenician "aleph" ("ox") and "beth" ("house") which were once pictograms. So, incredibly, the word "alphabet" contains within it the whole history of this particular alphabet or "ox-house."

Lloyd Daly (1967), author of the most in-depth study of the history of alphabetization to date, notes that the practice became possible when the ancient Greeks adopted the Phoenician alphabet, along with its established letter order. Yet, for roughly five centuries, the Greeks found no need to develop alphabetization, relying instead on other forms of classification, or indeed no classification scheme at all, to compile their lists. Daly traces the earliest uses of alphabetization to the end of the third century BCE. On the islands of Kos and Kalymnos, he finds inscriptions recording participants in local cults in which individuals' names were divided into sections and then arranged according to their first letter. The Alexandrian libraries provided an early occasion to apply the alphabetic principle more broadly, as scholars accumulated and needed to navigate amongst an expanding number of texts. Portions of the *Pinakes*, Callimachus' partial library catalogue, classified works by subject and then, most likely, by author. As part of their literary study, scholars also produced glosses of words found in various texts. At first, they arranged these lists to reflect the order in which the terms appeared in a given work, but as the glosses grew to unwieldy proportions, they began to arrange them alphabetically by first letter.

In spite of these early examples, Daly stresses that adoption of alphabetic order was piecemeal, and favored mainly by scholars rather than public officials. Although he finds evidence that tax rolls and other administrative documents from the Ptolemaic and Graeco-Roman administration of Egypt reflected alphabetization to some extent, he also explains that "for each example cited, there are hundreds of documents where the principle might have been used but was not" (50). One particular gap is found in the administration of ancient Rome, where the alphabetic scheme, although known, was not adopted to organize army rolls or tax ledgers, whose large scale might have benefited from such a system.

2.1 Some challenges of alphabetization

In all the early instances uncovered by Daly, alphabetization was limited to arranging items based on their first letter. Eventually, scholars began to extend the practice to order entries according to their second and third letters, but it is not until the second century CE that Daly finds examples of exact or absolute alphabetical order in Galen's *Interpretation of Hippocratic Glosses*.[10] In general, its cumbersome nature prevented absolute order from gaining widespread acceptance until the end of the Middle Ages, in part due to the effort required and in part due to the availability of materials. When compiling a list, an alphabetizer needed to estimate ahead of time the amount of area required to accommodate the number of entries falling under a given letter. Expanding or combining lists thus required physically fitting new items into pre-allotted spaces, which sometimes resulted in creating sub-lists or squeezing new entries into the margins of existing documents.

Until the development of printing, the alphabetic principle was also limited by media. Extensive alphabetization projects depend upon the ability to manipulate entries individually, and this is often done by first composing these entries on provisional cards[11] or slips. Both papyrus and parchment were too valuable to be used so ephemerally, and so until paper became cheaper and more abundant in the late fifteenth century, few efforts were made to apply alphabetization to its full potential. As Geoffrey Martin (2003, 16) writes, alphabetical indexes based on absolute order became common only "as the printed book established itself as an engine of scholarship," and alphabetization "came into its own as a guide to the contents of the greatly expanded libraries that printing made both possible and necessary to the advancement of learning."

Even in the age of the printed book, however, alphabetization remained one of many arrangement possibilities, and end users still needed to be guided in its application. When Robert Cawdrey published one of the first English dictionaries in 1604, his *Table Alphabeticall*, he explicitly instructed readers how to use it (quoted in Daly 1967, 91):

> If thou be desirous… rightly and readily to understand, and to profit by this Table, and such like, then thou must learne the alphabet, to wit the order of the letters as they stand, perfectly without booke, and where every letter standeth: as (b) nere the beginning, (n) about the middest, and (t) toward the end.

Clearly, Cawdrey could not assume that his early seventeenth-century readers were familiar with the practice of locating information by consulting alphabetically arranged documents.

There is also a point to be made about alphabetical order in "word" books at a time when spelling was not standardized and much more fluid. Mulcaster's *Elementarie* (1582) provides an example (http://www.bl.uk/learning/images/texts/dict/large1323.html). Words such as "chalenge," "chauffinch," "chearfull," and "chearie" are spelled differently in modern English, and, therefore, fall in different places in the alphabetic sequence.

Writers and publishers of encyclopedias have also wrestled with presenting alphabetic schemes to their readers. Etymologically, encyclopedias offer "general education" or "instruction in a circle," and most early authors sought to structure their works in ways that presented a coherent sum of human knowledge, stressing the internal relations between different fields of inquiry. Alphabetical arrange-

ments, in contrast, disperse conceptually related terms based on the relative happenstance of the order of their letters, severing important connections between associated ideas. Richard Yeo (1991) has written about this tension and the ways that, at least since the 1728 release of Ephraim Chambers' *Cyclopaedia*, editors have tried to resolve it using tools like subject indexes, cross-references, mixed thematic and alphabetical arrangements, and historical surveys. Chambers opted to combine systematic and alphabetical orders in his work, while the *Encyclopaedia Britannica*, in 1824, introduced longer historical dissertations on different branches of science to accompany its shorter, alphabetically-ordered entries.

The apparent objectivity of alphabetical order also obscures editorial decisions, such as whether one term should fall under the umbrella of another or merits its own treatment. Other practical concerns arise when a recently published volume contains entries that rely on concepts that follow them alphabetically and may not appear in print for years to come. Later entries might also be condensed to meet publication deadlines, space limits, and financial constraints. Yeo (40) points out that in the original *Encyclopaedia Britannica*, volumes dedicated to the letters A and B were granted 687 pages of text, while the remainder of the alphabet was condensed to occupy only 2,000 pages. Rather than necessarily offering order and ease of use, then, alphabetization is also capable of producing disorder, randomness, and opacity.

3.0 Some principles of alphabetization

Any arrangement scheme must take all elements of an index entry into consideration. Wellisch (1999) provides a detailed discussion of alphabetical arrangement and presents the following seven rules for ordering characters:

1. Headings shall be arranged exactly as written, printed or otherwise displayed. The arrangement of a heading among other headings should be based solely on the sequence of numbers in arithmetical order and on the sequence of the twenty-six letters of the English alphabet.
   The basic order of characters should be in the following sequence:
   – Spaces
   – Symbols other than numerals, letters and punctuation marks
   – Numerals (0 through 9)
   – Letters (A through Z)
2. Qualifying or explanatory terms are integral parts of a heading and should be arranged as any other words in the heading.
3. Headings beginning with identical words should be arranged in the following sequence. First: Single-word headings; Second: Multi-word headings, including headings with qualifiers.
4. Cross-references are not part of a heading, and therefore do not affect the arrangement of a heading.
5. Subheadings are normally arranged in alphanumeric sequence. They are subject to the same arrangement rules as the headings they modify. Function words at the beginning of subject headings should be arranged as any other words. They should not be disregarded.
6. An initial article in a heading should be treated as any other initial word. When it is deemed appropriate or desirable to arrange headings with initial articles by the word following the article (for example, in library catalogs where many title headings begin with an article) the heading may be structured to achieve the desired arrangement. Such structuring has two disadvantages: (a) it needs human intervention; and (b) the deletion of an article may distort the meaning of a heading, especially in titles.
7. Numbers in headings, whether at the beginning or within a heading, should be arranged in arithmetic order. Headings beginning with numbers written in Arabic numerals should be sorted in ascending arithmetic order before headings beginning with a letter sequence. Roman numbers (written by means of letters) should be arranged by their arithmetical value, among other numbers written in Arabic numerals. To achieve this, the sequence of letters must first be tagged as a number by human intervention, and may then be sorted as a Roman numeral, either manually or by an algorithm.

There are two overall basic forms of arrangements of headings, word-by-word and letter-by-letter (see table 1). These two schemes differ in how they handle spaces and other non-letter characters and typographical forms. Word-by-word arrangement puts "nothing before something," whereas letter-by-letter arrangement ("all through") ignores spaces and punctuation between words. Wellisch (1999, 5 emphasis original) writes:

> This method [letter-by-letter] is primarily used for the arrangement of headings in dictionaries, because it keeps different spellings of the same term together (for example, ground water, ground-water, groundwater). The application of this method violates, however, the provision of Section 3.1, and it is also

subject to a number of different interpretations …. This method is therefore *not recommended*."

As shown in Table 1, these two styles of alphabetizing yield very different results, which is of great importance in long indexes. Most users of indexes do not think about the various ways entries may be alphabetized, and if not found in a particular place, they may assume that a subject is not included. Using standards and orders that work for users is critical. Unfortunately, there are different, non-compatible standards and guidelines, as discussed below.

4.0 Standards, norms and guidelines[12]

The rules and standards governing the many aspects of alphabetization may be difficult to grasp. This is particularly true with the implementation of well-established national traditions for arranging names and headings in computer programs. This process is often (e.g., in *Library of Congress Filing Rules* as well as in this article) called "filing rules" (see note thirteen about the use of this term in classification research).

One challenge of alphabetization in computer software is establishing the method by which a system will encode alphanumeric characters. The encoding of characters in computer systems has been guided by both national and international standards, as well as by proprietary encoding schemes established by the various software houses, e.g., IBM, Microsoft, Apple Computer etc., leading to difficulties in interoperability between different software programs.

As an example, uppercase and lowercase letters are ordered separately following the 7-bit character set defined by the *American Standard Code for Information Interchange* (ASCII) (Table 2; for ASCII and bit see Appendix 1 and 2). This does not follow the traditional ordering of letters in the English alphabet, where uppercase and lowercase letters do have the same position in the alphabet.[14] In a digital computer (or binary computer), each character is given a unique binary code. This means that an uppercase A has a different code than a lowercase a. According to ASCII, all uppercase letters appear in order first, followed by lowercase letters. Following this logic, all entries beginning with uppercase letters will be arranged before entries beginning with lowercase letters.

Table 2 illustrates the result of using the ASCII arrangement to encode characters compared to the example in Table 1.

In the two leftmost columns, the arrangement follows the traditional English alphabetical order according to the guiding principles of word-by-word or letter-by-letter arrangements, with no distinction between uppercase and lowercase letters. In the rightmost column, the order follows the encoding scheme used in the ASCII character set.

In addition to encoding schemes, it has, therefore, been necessary to establish guides or collation[15] rules for how letters should be ordered according to national alphabets. These language-specific rules reflect different cultural traditions for arranging alphabetic characters.

To add to the complexity, different institutions (e.g., libraries, publishing houses) also maintain specific traditions for how they arrange names and headings. This impacts the order of books on shelves, the arrangement of book indexes, and the display of search results in an OPAC.

Ordering practices have been guided by professional associations like the American Library Association, the Library of Congress, ARMA International (previously the

Word-by-Word	Letter-by-Letter (Strict interpretation)
N. E. Zenith Co.	networks
networks	New, Agnes
New, Agnes	New Brunswick
New Brunswick	N. E. Zenith Co.

Table 1. Simplified figure after Wellisch (1999, 6)

Word-by-Word	Letter-by-Letter (Strict interpretation)	ASCII (unmodified)
N. E. Zenith Co.	networks	"New lamps for old"
networks	New, Agnes	N. E. Zenith Co.
New, Agnes	New Brunswick	New Brunswick
New Brunswick	N. E. Zenith Co.	networks

Table 2. Simplified figure after Wellisch (1999, 6).

Association of Records Managers and Administrators), as well as by standardizing bodies such as the National Information Standards Organization (NISO) and International Organization for Standardization (ISO), among others. Filing rules differ by the level of human intervention used to determine which part of the heading or name should be used for ordering. This involves an intellectual understanding of the actual meaning of the heading, i.e., to distinguish between a personal name, a place name, a subject etc. and arrange accordingly. The example below is taken form the *Library of Congress Filing Rules* (1980, 24), where headings with identical leading elements[16] are arranged in the following order: person, place, corporate body, subject, title (leading element underlined):

George III, King of Great Britain, 1738-1820
George, Saint, d. 303
George, Alan
George (Ariz.)
George (Motor boat) [corporate body]
George, Lake, Battle of, 1755 [subject heading]
George [motion picture]

In this example, the leading element is in all cases identical and the list is then arranged according to type of heading.

Outside the scope of this article are the standards, rules, and guidelines suggesting what indexes are appreciated in a certain document or information system and how entries or headings should be formulated, e.g., back-of-the-book indexes, algorithmic search indexes, library OPACs, etc.

Figure 1 provides an overview of numerous standards, guidelines, and rules (the top box represents issues related to indexing, cataloging, and metadata that are beyond the scope of the present article[17]). In Figure 1, 4.1 and 4.2 denote standards guiding the encoding of characters and technical solution for the implementation of filing rules in computer systems. 4.3 is filing rules and the order of letters used within different domains. 4.1 to 4.3 are explained in detail below. The dotted box gives examples of guidelines and rules for the formulation of headings and indexes etc. "AACR 2ed." is the Anglo American Cataloging Rules and is probably the most widely used cataloguing rules globally. "RDA" is The Resource Description and Access cataloguing standard (Joint Steering Committee for RDA 2015) and is considered the successor to AACR2. These guidelines fall outside the scope of this article, but all mentioned standards and guidelines are included in the reference list.

4.1 Standards for encoding of alphanumerical characters

Presented below is a selection of US and international standards, mainly governing the encoding of the English written alphabet with later extensions allowing for encoding of alphabets using Latin script.

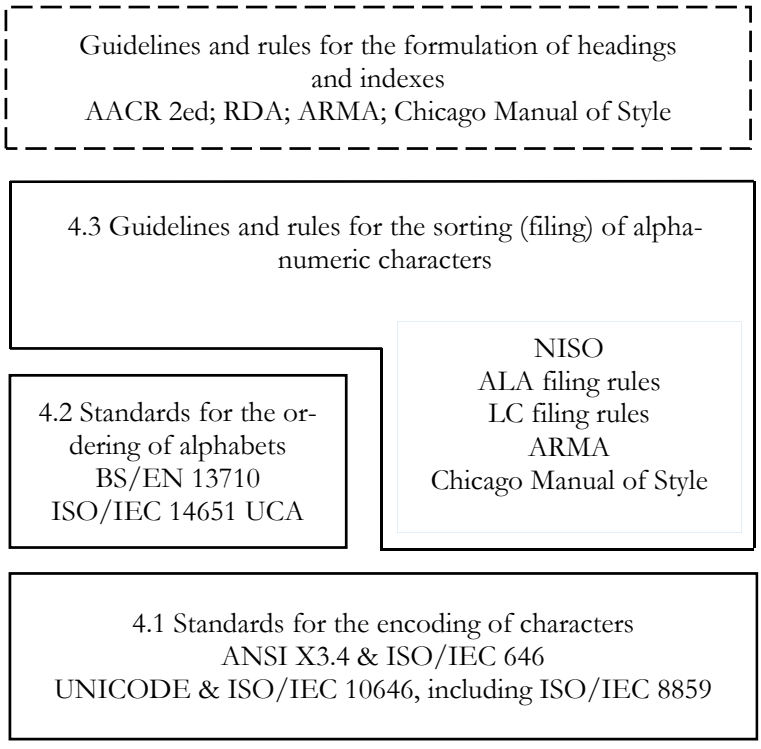

Figure 1. Overview of standards, guidelines and rule.

ANSI INCITIS X3.4-1986: Information Systems—Coded Character Sets—7-Bit American National Standard Code for Information Interchange (7-bit ASCII), first edition published in 1963 and was adapted as the international standard ISO/IEC 646 in 1967. These two standards for 7-bit encoding are only presented here because of their historical importance for the early development of computers and the attempt to standardize the industry. The 7-bit character sets provided space for English alphanumeric letters, resulting in many national variants. To support a wider number of characters, the 8-bit family of encoding standards was developed, the first edition published in 1987 as ISO/IEC 8859. This family of standards is incorporated in ISO/IEC 10646 mentioned below.

A widely used character set is the *Unicode Standard*, which was first published in 1991 and whose most recent version, Unicode 11.0, was published in 2018 (Unicode Consortium 2018). Version 11.0 contains a repertoire of 137,439 characters covering 146 modern and historic scripts, as well as multiple symbol sets and emoji.[18]

Unicode makes it possible to encode more than 1.1 million characters, thereby providing encoding of all existing alphabets, including letter based as well as ideographic writing systems, but only a fraction of this set is currently in use. The Unicode standard is developed by The Unicode Consortium in tandem with ISO, and the most recent ISO standard is ISO/IEC 10646:2017 Information technology—Universal Coded Character Set (UCS). It corresponds to Unicode 10.0 but excludes some special characters and emoji symbols (See further Wikipedia Universal Coded Character Set 2018).

Unicode is currently the most important issue relating to alphabetization, and it may deserve an independent entry in this encyclopedia (see Aliprand 2017 for an encyclopedia article in *Encyclopedia of Library and Information Sciences*). From a research-oriented perspective, two issues are crucial: 1) unicode can be implemented by different character encodings and there seems to be a trade-off between the number of bytes used for each character and the space used, and thus the efficiency of the implementation; and, 2) philosophical and completeness criticisms. There has been a debate on such issues (see endnote nineteen). Among the issues raised is the relation between characters, graphemes and glyphs as units. Holmes (2003) has suggested that although Unicode is a success, a different approach would have worked much better for encoding text, documents and writing systems. The attempt to accommodate all the world's languages in one gigantic codespace means that it cannot take full advantage of the systematic graphical features of various writing systems. The criticisms of Unicode seem related to earlier versions and are possibly less relevant to its newer versions. It is, however, important to be open to possible limitations and biases in all kinds of standards and knowledge organization systems.

4.2 Standards and recommendations for the ordering of alphabets

According to Küster (1999, 21) the "ordering of letters is highly dependent on the cultural expectations." This author thus seems to strive for a multilingual approach to ordering. What might be expected as the correct alphabetical order in English is not the same in, for example, Danish.

Besides the letters a to z, the Danish alphabet also comprises the letters æ, ø and å, and the ordering of the Danish alphabet is from a – å, meaning that æ, ø and å are the three last letters in the alphabet. This raises a number of questions about how to treat different national alphabets when dealing multilingual information and software. These issues are both about securing correct order according to different national traditions and about how to incorporate or express letters from other alphabets in, for example, the English language.

Example: according to Wellisch (1999, 3) the Danish letters *æ, ø and å* should be arranged in the English alphabet as ae, o and a. Needless to say, this would have an effect on the arrangement of characters when following a Danish language-based system compared to an English language-based system, and subsequently also the exchange of information between the two systems. This is not just of "academic interest" but relevant whenever Danish names appear in English reference lists—and of course similarly with every other language.

Standards such as BS/EN 13710: 2011 *European Ordering Rules. Ordering of Characters from Latin, Greek, Cyrillic, Georgian and Armenian Scripts* have been established to normalize this (see also Küster 2006, chapter 17.4).

The standards and recommendations mentioned here do not only deal with the ordering of letters but also define collation algorithms. According to Davis, Whistler and Scherer (2018, Section 1) the purpose of the Unicode Collation Algorithm (UCA) is:

> Collation varies according to language and culture: Germans, French and Swedes sort the same characters differently. It may also vary by specific application: even within the same language, dictionaries may sort differently than phonebooks or book indices. For non-alphabetic scripts such as East Asian ideographs, collation can be either phonetic or based on the appearance of the character. Collation can also be customized according to user preference, such as ignoring punctuation or not, putting uppercase before lowercase (or vice versa), and so on. Linguistically correct searching needs to use the same mech-

> anisms: just as "v" and "w" traditionally sort as if they were the same base letter in Swedish, a loose search should pick up words with either one of them.

Collation rules have a wide impact on digital systems, from determining the simple alphabetical ordering of letters in an index to influencing how databases and search engines are organized and consequently behave when confronted with a search request submitted by a user.

One important function of UCA is, therefore, to provide a technical solution for implementing filing rules (see below in 4.3) in a software program. It is imperative to underline that the collation algorithm does not prescribe specific rules for how to arrange or file headings; it only governs the technical implementation of filing rules.

The international collation standard is ISO/IEC 14651, *Information Technology, International String Ordering and Comparison, Method for Comparing Character Strings and Description of the Common Template Tailorable Ordering*. It was developed in tandem with UCA. Furthermore, Wellisch (1999) and the *LC Filing Rules* (1980) prescribe the ordering of the English alphabet and the arrangement of non-English letters into the English alphabet.

4.3 Rules and guidelines for the arrangement of headings (filing rules)

Filing rules guide alphabetization, including the ordering and sorting of library catalogs, indexes, dictionaries, and directories (Wellisch 1999, v). These rules are published by both professional entities and organizations, e.g., national library bodies, library associations, publishing houses, etc. With this in mind, only a few important examples of guidelines are mentioned here.

Wellisch (1999) published by NISO is an attempt to establish a set of common guidelines. According to the foreword, "this technical report seeks to make the alphanumeric arrangement of headings 'as easy as ABC'" (Wellisch 1999, v). The American Library Association (ALA) has published *ALA Filing Rules* (American Library Association 1980) and the Library of Congress has published *LC Filing Rules* (Rather and Biebel 1980). Both are widely used within libraries, but alas they provide different solutions. For example, the *ALA Filing Rules* do not distinguish between types of headings (Bakewell 1972, 166); this differs from the *LC Filing Rules* (see this article Section 4 for example). The three recommendations above all advise a word-by-word arrangement.

Many book publishers follow their own alphabetizing styles. North American publishers often follow the guidelines in *The Chicago Manual of Style* (University of Chicago Press 2017, 944, §16.58), which call for letter-by letter: Alphabetization: "Chicago, most university presses, and many other publishers have traditionally preferred the letter-by-letter system but will normally not impose it on a well-prepared index that has been arranged word by word." It is important to note that Chicago's choice of letter-by-letter alphabetization is in conflict with the word-by-word arrangement recommended by Wellisch (1999) and by both the *ALA Filing Rules* and the *Library of Congress Filing Rules*.

For use within the domain of Records and Information Management, ARMA International (ANSI 2005) publishes a set of guidelines. These guidelines advise a unit-by-unit approach for alphabetical filing, which differs from both letter-by-letter and word-by-word filing.

5.0 Alphabetic order versus other ordering criteria

In botany, Richards (2016, 66) explains that alphabetical arrangements of plants in herbaria were common by about 1596, but many other criteria were also used, like sorting plants with pleasant flowers from odorous plants and classifying plants according to their similarities and differences. This last principle led to hierarchical and more systematic approaches, for instance, organizing plants into genera and subdividing them into species. But these species and genera were not necessarily what we would see in modern scientific classifications. Sometimes plants were, for example, simply classified as trees, shrubs, or herbs. It is common knowledge that such different ordering principles were standardized by the taxonomy set up by Carl Linnaeus in his *Systema Naturae* (1735). Today it is the norm that such systematical arrangements are supplemented by alphabetic indexes for the easy location of a specific name.

Concerning the organization of knowledge in encyclopedias, Sundin and Haider (2013) write:

> The encyclopaedias that emerged around the time of the Enlightenment are said to have shifted knowledge's organizational principle; from the tree of knowledge to the alphabet. Yet despite the success of the alphabetic principle, it has not erased classification endeavours, in fact not even in the beginning. As Ann Blair [2010] points out, already d'Alembert defended the alphabetic principles in the Encyclopédie at the same time that he provided readers with an image of a tree of knowledge as a supplement to the alphabet.

Sundin and Haider then describe how the Swedish electronic encyclopedia *Nationalencyklopedin,* in addition to its alphabetical arrangement, also uses a Swedish bibliographic classification system *Klassifikationssystem för svenska bibliotek* (*SAB*). However, the authors do not further examine the use of classification systems in contemporary encyclope-

dias, and although such systems are sometimes provided (e.g., in *Encyclopedia Britannica*'s "Syntopicon: An Index to The Great Ideas" (1952) followed in 1974 by *Propaedia*, an "outline of knowledge," see Adler 2007), there is little evidence of their use and usefulness over alphabetical arrangements, indexes and internal references. However, such systematic outlines often form the basis for the overall editing of encyclopedias and the commission of articles. For the user, they may, therefore, provide a better overview and means to evaluate the coverage of the work.

In libraries, there have been controversies about the strengths and weaknesses of alphabetical subject catalogs versus systematic catalogs (see Hanson and Daily 1970 about the history of library catalogs). In *The Organization of Knowledge in Libraries and the Subject-Approach to Books*, Henry E. Bliss (1933) argues that a systematic subject-approach is required. Any attempt to apply a simple alphabetical subject-approach without a systematic organization of the plurality of knowledge subjects is rejected by Bliss (1933, 301) as a kind of "subject-index illusion."[20] A mere listing of subjects, as provided by subject headings, would not be able to meet the principle of maximal efficiency that results from the strategies of collocation of closely related classes or subjects and subordination of the specific to the generic. This means that a differentiation (analysis) of subjects should only be considered as a necessary first step that needs to be succeeded by an integration (synthesis) of subjects into a well-structured knowledge organization system, as underlined by Bliss (1933, 104):

> Analytic division tends to dispersion. But synthesis, either collocative or systematic, places subjects in effectual relation and efficient organization. A collocative synthesis does not, however, forego analysis, which inevitably issues from subdivision; but it collocates the results of analytic subdivision. This is the very nature of systematic classification. It opposes the false theory that disorder and dispersion can be obviated or compensated by an alphabetic key or subject-index.

There are different ways of combining alphabetic and systematic order. One example is provided by the so-called "Cutter numbers" used by the Library of Congress, where alphabetic arrangement is a very significant aspect of the classification scheme.[21]

6.0 Conclusion

Research has demonstrated the complexities that may arise from using alphabetization: the apparent simple process can be quite difficult. To order headings and indexes alphabetically is not as straightforward as it may sound, depending on both cultural traditions and different approaches used in different domains or under different circumstances. The implementation of well-established filing rules in computer software has resulted in a number of different proprietary technical solutions established by software companies. What has characterized these has been a lack of interoperability, resulting in incompatible systems. The development of computers and software has been dominated by Anglo-American companies; hence, the default "computer" language has been and still is English. This has created a number of difficulties for supporting non-English alphabets, based on both Latin and non-Latin writing systems. Fortunately, the increase in computational power and decrease in storage cost has led to the development of new standards like Unicode, which can support all known writing systems. Unicode has now gained ground as the "default" standard for encoding characters, compatible with virtually all modern computer software. It now seems possible to support our culturally diverse writing systems and to achieve interoperability between different computer software. However, technical as well as philosophical questions persist: What happens when the most comprehensive standards prove impractical to use? And can any alphabetization standard ever function as a neutral tool, or will it always serve some cultures and domains better than others?

Notes

1. This entry is about written alphabets only. We are not addressing issues relating to unwritten languages or sign-languages. About the International Phonetic Alphabet see Brown (2013).
2. This first meaning of ordering corresponds to how WordNet 3.1. defines the noun "ordering": S: (n) order, ordering (the act of putting things in a sequential arrangement) 'there were mistakes in the ordering of items on the list'." Küster (1999, 21; italics in original) made a distinction between sorting and ordering that conflicts with the other definitions presented here: "English terminology usually distinguishes between sorting and ordering. Sorting is primarily a service for users to facilitate their access to information by presenting it in a structured and predictable way, e. g. by subdividing the information by subject matter (by having several registers to a book, for instance), having multiple indices in a library etc. Ordering—the arrangement of information in alphabetical sequence—is in most cases an integral part of this procedure." But as we saw the term ordering is not normally limited to alphabetization.
3. Even a random order may be used for some purposes, e.g., statistical sampling.

4. The *Oxford English Dictionary* (2018) defines sorting: "9. a. transitive. To arrange (things, etc.) according to kind or quality, or after some settled order or system; to separate and put into different sorts or classes; to classify; to assort."
   WordNet 3.1 defines sorting (as a noun):

   – S: (n) sort, sorting (an operation that segregates items into groups according to a specified criterion) "the bottleneck in mail delivery is the process of sorting"
   – S: (n) classification, categorization, categorisation, sorting (the basic cognitive process of arranging into classes or categories)
   – S: (n) sorting (grouping by class or kind or size)

   ODLIS: Online Dictionary for Library and Information Science (Reitz 2004) defines sorting: "In a search of a online catalog or bibliographic database, the default display is normally alphabetical order by author or title, or reverse chronological order by publication date. However, in some online catalogs and databases, the user may select the sequence in which results will be displayed, usually from a list of options, either before or after the search is executed. Compare with ranking. See also: arrangement."
5. About algorithmic sorting see, for example, Knuth (1998), Christophersen (1997) and Wikipedia: "Sorting Algorithm" at https://en.wikipedia.org/wiki/Sorting_algorithm.
6. Wellisch (1999, 2) defines heading: "Any written, printed or otherwise visually displayed item, consisting of one or more words, that is to be arranged among other such items in a known order."
7. A character is the "smallest possible unit of arrangement: a space, letter, numeral, punctuation mark, or other symbol" (Wellisch 1999, 1). Later, in Section 4.1) it is mentioned that the UNICODE has met some difficulties with characters and that glyphs rather than characters may be needed as units in some scripts.
8. In practice, library catalogs will mostly apply the principle of uniform titles to ensure that a translation is entered under the original title to keep versions of the same work together.
9. One of the anonymous referees wrote: "Otherwise I thought this was a firm rebuttal of Weinberger and a challenge to the idea that alphabetical order is arbitrary, on that basis almost every ordering principle is, and even 'natural' orders need to seek consensus on the sequence (e.g. natural numbers in ascending order, elements in the periodic table by increasing atomic number and weight). What is a 'natural' order (such as the elements) may not be familiar to a lay audience in the manner of alphabetical order, and hence completely ineffective for retrieval."
10. Valerius Harpocration was, according to Keaney (1973) probably the first to use absolute alphabetization.
11. In this context, it seems relevant to mention that it was Carl Linnaeus (1707–1778) who invented the card index (cf., Mueller-Wille 2009). The card index served an important purpose: "Linnaeus had to manage a conflict between the need to bring information into a fixed order for purposes of later retrieval, and the need to permanently integrate new information into that order."
12. Beside the guidelines mentioned in the section, Chauvin (1977) should be mentioned.
13. In classification research, in particular in the facet-analytic tradition, the terms "citation order" and "filing order" are well established with the following meanings:

    – "Citation order simply refers to the order in which notational elements are cited in a built notation. The most commonly applied rule is to cite the most specific concept first and then move in stages to the most general" (Batley 2005, 17).
    – "Filing order, which establishes shelf order, is usually the opposite of the citation order, with general aspects of a subject shelved before more specific aspects. This makes intuitive sense: library users would expect broad aspects of a topic to be shelved before narrower aspects" (Batley 2005, 17).

    In a printed telephone directory, for example, this terminology implies that the citation order would be the way the single entry is constructed (e.g., "Adams, John W. librarian #xx) while the filing order would be the alphabetical arrangement of the different entries.
14. ASCII-code order is also called ASCIIbetical order. In ASCII all uppercase come before lowercase letters; for example, Z precedes a (see the ASCII table in Appendix A).
15. See also "Collation" in Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Collation
16. Headings are split into elements where an element can consist of one or more words and is identified by punctuation marks etc., e.g., a person's name consisting of a last name, first name is split in two elements using the comma as delimiter.
17. The history of the *AACR* cataloging rules and the different editions can be seen in Joint Steering Committee for RDA (2009) http://www.rda-jsc.org/archivedsite/history.html; The latest version of the RDA is published by Joint Steering Committee for RDA (2015). Such rules belong to the field of (descriptive) cataloging (see

Joudrey 2017). Publishers' guidelines (Such as the *Chicago Manual of Style* (University of Chicago Press 2017)) are mainly constructed from practical experience, but there is a growing tendency to consider normative guidelines from the perspective of genre- and writing studies, thus contributing theoretical perspectives.

18. Most editions are published in electronic format as well as book form and have an ISBN; however, newer editions are not available in WorldCat or in Amazon.com but a pdf can be generated from the unicode.org page. Details about the book publication and ordering information of Unicode standards may be found at http://www.unicode.org/book/aboutbook.html
19. A debate included Goundry (2001) "Why Unicode Won't Work on the Internet:
Linguistic, Political, and Technical Limitations"; Whistler (2001), "Why Unicode Will Work On The Internet." Peterson (2006) "Unicode in Japan. Guide to a Technical and Psychological Struggle and Searle (2002)." Unicode Revisited. " There is also in Wikipedia's entry about Unicode section about this: https://en.wikipedia.org/wiki/Unicode#Philosophical_and_completeness_criticisms
20. However, despite Bliss' criticism, the dictionary catalog had many followers, and there was a good deal of opposition to his view, most notably by John Metcalfe (1959).
21. Named after Charles Ammi Cutter, Cutter numbers represent a method of representing words or names by using a decimal point followed first by a letter of the alphabet, then by one or more Arabic numerals. In Library of Congress (LC) Call numbers, Cutter numbers do function as book number and distinguishes a particular work from others in the same class. "Example: Call number: Z733.U58G66 1991 contains Cutter number: .U58 [for the United States] and G66 [for Goodrum, the author]." (Example taken from https://www.itsmarc.com/crs/mergedProjects/cutter/cutter/definition_cutter_number_cutter.htm). Immroth (1971, 384) wrote that Cutter "is perhaps best known today for his alphabetic order of Cutter tables." Winke (2002) provides an overview of the current use of Cutter's Expansive Classification of which only Cutter numbers and Cutter tables remains in general use.

References

Adler, Mortimer J. 2007. "Circle of Learning." *The New Encyclopædia Britannica*, 15th ed. Chicago: Encyclopædia Britannica.

Aliprand, Joan M. 2017. "Unicode Standard." In *Encyclopedia of Library and Information Sciences*, 4th ed., ed. John D. McDonald and Michael Levine-Clark. Boca Raton, FL: CRC Press, 7: 4662-71.

American Library Association. 1980. *ALA Filing Rules*. Chicago: American Library Association.

ANSI (American National Standards Institute.) 1986. *American National Standard for Information Systems. Coded Character Sets. 7-bit American National Standard Code for Information Interchange (7-Bit ASCII)*. New York: American National Standards Institute.

ANSI (American National Standards Institute). 2005. *Establishing Alphabetic, Numeric and Subject Filing Systems*. Lenexa, KS: ARMA International.

Bakewell, Kenneth Graham Bartlett. 1972. *A Manual of Cataloguing Practice*. Oxford: Pergamon.

Batley, Sue. 2005. *Classification in Theory and Practice*. Oxford: Chandos.

Blair, Anne. 2010. *Too Much to Know: Managing Scholarly Information before the Modern Age*. New Haven, CT: Yale University Press.

Bliss, Henry Evelyn. 1933. *The Organization of Knowledge in Libraries and the Subject-Approach to Books*. New York: Wilson.

Brown, Adam. 2013. "International Phonetic Alphabet." In *The Encyclopedia of Applied Linguistics*, ed. Carol A. Chapelle. 10 vols. Hoboken, NJ: Wiley-Blackwell, 5: 1-8. DOI: 10.1002/9781405198431.wbeal0565

BS/EN 13710: 2011. *European Ordering Rules. Ordering of characters from Latin, Greek, Cyrillic, Georgian and Armenian scripts*. London: British Standards Institution.

Chauvin, Yvonne. 1977. *Pratique du Classement Alphabétique*. 4th ed. Paris: Bordas.

Christophersen, Hans 1997. "Alphabetisierung auf Computer. Prinzipien, Probleme und eine Lösungsverbesserung." http://www.rostra.dk/alphabet/alpha%5Fdt.htm

Daly, Lloyd W. 1967. *Contributions to a History of Alphabetization in Antiquity and the Middle Ages*. Brussels: Latomus.

Daniels, Peter T. and William Bright, eds. 1996. *The World's Writing Systems*. New York: Oxford University Press.

Davis, Mark, Ken Whistler and Markus Scherer. 2018. "Unicode Collation Algorithm (11.0.0)." Unicode Technical Standard 10. https://www.unicode.org/reports/tr10/

Diringer, David. 1962. *Writing*. London: Thames & Hudson.

Drucker, Johanna. 1995. *The Alphabetic Labyrinth: The Letters in History and Imagination*. London: Thames & Hudson.

Goundry, Norman. 2001. "Why Unicode Won't Work on the Internet: Linguistic, Political, and Technical Limitations." Technical Papers. http://www.hastingsresearch.com/net/04-unicode-limitations.shtml

Hanson, Eugene R. and Jay E. Daily. 1970. "Catalogs and Cataloging." In *Encyclopedia of Library and Information Science*, ed. Allen Kent and Harold Lancour. New York: Marcel Dekker, 4: 242-305.

Holmes, Neville. 2003. "The Problem with Unicode." *Computer* 36, no. 6: 116 + 114-115 [sic!]. doi:10.1109/MC.2003.1204385

Hooker, James T., ed. 1990. *Reading the Past: Ancient Writing from Cuneiform to the Alphabet*. London: British Museum Press.

Immroth, John Phillip. 1971. "Cutter, Charles Ammi." In *Encyclopedia of Library and Information Science*, ed. Allen Kent and Harold Lancour. New York, NY: Marcel Dekker, vol. 6: 380-7.

ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission). 1991. *Information Technology, ISO 7-Bit Coded Character Set for Information Interchange*. International Standard ISO/IEC 646. Geneva: International Organization for Standardization and International Electrotechnical Commission.

ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission). 1999. *Information technology, 8-Bit Single-Byte Coded Graphic Character Sets*. International Standard ISO/IEC 8859. Geneva: International Organization for Standardization and International Electrotechnical Commission.

ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission). 2017. *Information Technology: Universal Coded Character Set (UCS)*. 5th ed. International Standard ISO/IEC 10646. Geneva: International Organization for Standardization and International Electrotechnical Commission. Reference number: ISO/IEC 10646:2017(E).

ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission). 2016. *Information Technology, International String Ordering and Comparison, Method for Comparing Character Strings and Description of the Common Template Tailorable Ordering*. 4th ed. International Standard ISO/IEC 10646. Geneva: International Organization for Standardization and International Electrotechnical Commission. Reference number: ISO/IEC 10646:2017(E).

Joint Steering Committee for RDA. 2009. "A Brief History of AACR." http://www.rda-jsc.org/history.html

Joint Steering Committee for RDA. 2015. *RDA: Resource Description and Access*. 2015 Revision. Chicago: American Library Assn.

Joudrey, Daniel N. 2017. "Cataloging." *Encyclopedia of Library and Information Sciences*, 4th ed., ed. John D. McDonald and Michael Levine-Clark. Boca Raton, FL: CRC Press, 2: 723-32.

Keany, John. J. 1973. "Alphabetization in Harpocration's Lexicon." *Greek, Roman, and Byzantine Studies* 14: 415–23.

Knuth, Donald E. 1998. *Sorting and Searching*. Vol. 3 of *The Art of Computer Programming*. 2nd ed. Boston: Addison-Wesley.

Küster, Marc Wilhelm. 1999. "Multilingual Ordering: The European Ordering Rules." In *Multilinguale Corpora. Codierung, Strukturierung, Analyse. 11. Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung*, ed. Jost Gippert and Peter Olivier. Prag: Enigma, 21–33.

Küster, Marc Wilhelm. 2006. *Geordnetes Weltbild. Die Tradition des alphabetischen Sortierens von der Keilschrift bis zur EDV. Eine Kulturgeschichte*. Berlin: De Gruyter.

Library of Congress. 1980. *Library of Congress Filing Rules*, prepared by John C. Rather and Susan C. Biebel. Washington: Cataloging Distribution Service, Library of Congress.

Lock, Andrew and Charles R. Peters, eds. 1999. *Handbook of Human Symbolic Evolution*. Oxford: Blackwell.

Mackenzie, Charles E. 1980. *Coded Character Sets: History and Development*. Reading, MA: Addison-Wesley Publishing.

Martin, Geoffrey. 2003. "Alphabetization Rules." In *International Encyclopedia of Information and Library Science*. 2nd ed., ed. John Feather and Paul Sturges. London: Routledge, 15-17.

Metcalfe, John. 1959. *Subject Classifying and Indexing in Libraries and Literature*. Sydney: Angus & Robertson.

Mueller-Wille, Staffan. 2009. "Carl Linnaeus Invented the Index Card." Paper presented at the annual meeting of the British Society for the History of Science in Leicester July 4, 2009. https://phys.org/news/2009-06-carl-linnaeus-index-card.html

Mulcaster, Richard. 1582. *The First Part of the Elementarie which Entreateth Chefelie of the Right Writing of our English Tung*. London: T. Vautroullier.

Oxford English Dictionary Online. 2018. S.v. "Ordering", accessed November 4.

Oxford English Dictionary Online. 2018. S.v. "Sorting", accessed November 4.

Peterson, Benjamin. 2006. "Unicode in Japan. Guide to a Technical and Psychological Struggle." Blog post. https://web.archive.org/web/20090627072117/http://www.jbrowse.com/text/unij.html

Reitz, Joan M. 2004. *ODLIS: Online Dictionary for Library and Information Science*. Santa Barbara, CA: Libraries Unlimited. https://www.abc-clio.com/ODLIS/odlis_a.aspx

Richards, Richard A. 2016. *Biological Classification: A Philosophical Introduction*. Cambridge: Cambridge University Press.

Rosen, Michael. 2015. *Alphabetical: How Every Letter Tells a Story*. Berkeley, CA: Counterpoint.

Searle, Steven J. 2002. "Unicode Revisited." http://tronweb.super-nova.co.jp/unicoderevisited.html

Sorting Algorithm. 2018. *Wikipedia*, accessed Nov. 3. https://en.wikipedia.org/wiki/Sorting_algorithm

Sundin, Olof and Jutta Haider. 2013. "The Networked Life of Professional Encyclopaedias: Quantification,

Tradition, and Trustworthiness." *First Monday* 18, no. 6. https://firstmonday.org/article/view/4383/3686

Unicode Consortium. 2018. *The Unicode Standard.* Version 11.0.0. Mountain View, CA: Unicode Consortium. http://www.unicode.org/versions/Unicode11.0.0/

University of Chicago Press. 2017. *The Chicago Manual of Style*, 17th ed. Chicago: University of Chicago Press.

Weinberger, David. 2007. *Everything Is Miscellaneous: The Power of the New Digital Disorder*. New York: Times Books.

Wellisch, Hans H. 1999. *Guidelines for the Alphabetical Arrangement of Letters and Sorting of Numerals and Other Symbols*. NISO Technical Report 1081-8006 vol. 3. Bethesda, MD: National Information Standards Organization.

Whistler, Ken. 2001. "Why Unicode Will Work on The Internet." *Slashdot* (blog), June 9. https://features.slashdot.org/story/01/06/06/0132203/why-unicode-will-work-on-the-internet

Unicode. 2018. *Wikipedia, a*ccessed Nov. 3. https://en.wikipedia.org/wiki/Unicode

Universal Coded Character Set. 2018. *Wikipedia,* accessed Nov. 3. https://en.wikipedia.org/wiki/Universal_Coded_Character_Set

Winke, R. Conrad. 2002. "The Contracting World of Cutter's Expansive Classification." *Library Resources & Technical Services* 48: 122-9.

Wordnet Search - 3.1. s.v. "ordering." Accessed November 4.

Wordnet Search - 3.1. s.v. "sorting." Accessed November 4.

Yeo, Richard. 1991. "Reading Encyclopedias: Science and the Organization of Knowledge in British Dictionaries of Arts and Sciences, 1730-1850." *Isis* 82, no. 1: 24-49.

Appendix 1: ASCII Table

ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	|
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Appendix 2

Developments in character codes by bits

- One-bit character sets can have two possible characters. 2^1=2. 0 or 1. (This is the binary alphabet used by modern computers)
- Two bits character sets can have four possible characters. 2^2=4. 00,01,10,11. (i.e. 0-3)
- Three bits character sets can have eight possible characters 2^3=8.
- Four bits character sets can have 16 possible characters. 2^4=16. 0000,0001,0010,0011, etc. (i.e. 0-15)

- Five bits character sets can have 32 possible characters. $2^5 = 32$. Until about 1928 some 5-bit codes were used (e.g., Baudot code and Murray code)
- Six bits character sets can have 64 possible characters. $2^6=64$. In 1928 the BCD ("Binary-Coded Decimal") 6 bits code was introduced with the IBM card, generally used for the upper-case letters, the numerals, some punctuation characters, and sometimes control characters.
- Seven bits character sets can have 128 possible characters. $2^7=128$. 0000000,0000001,0000010, etc. (i.e. 0-127). In 1963 the ASCII 7 bits code provides 128 different characters; in 1968 MARC-8 7 bits Library computer systems was introduced.
- Eight bits character sets can have 256 possible characters. $2^8=256$. 00000000,00000001,00000010, etc. (i.e. 0-255). In 1963 the Extended Binary Coded Decimal Interchange Code (EBCDIC) 8 bits code were developed for IBM computers.
- 16 bits character sets can have 2^{16} possible characters = 65,536
- 32 bits character sets can have 2^{32} possible characters = 4,294,967,296. In 1991 Unicode, packed into 8/16/32, but less than 21 bits are usable (=2,097,152 characters).