Engines of Order
Rieder, Bernhard

Published by Amsterdam University Press

Project MUSE. muse.jhu.edu/book/76715.

For additional information about this book
https://muse.jhu.edu/book/76715

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2652877
Part II
4. From Universal Classification to a Postcoordinated Universe

Abstract
This chapter starts off from standard takes on knowledge organization and classification in libraries and encyclopedias, but then zeros in on the field of information retrieval, which develops in fundamental opposition to even the most visionary of library techniques. Coordinate indexing, the first technique in this lineage, is explicitly designed to eliminate the influence of librarians and other knowledge mediators by shifting expressive power from the classification system to the query and, by extension, to the information seeker. Order is no longer understood as a stable map to the universe of knowledge but increasingly as the outcome of a dynamic and purpose-driven process of ordering. The chapter closes by discussing coordinate indexing as a precursor of the relational model for database management.

Keywords: bibliographic organization, coordinate indexing, information wars, relational model for database management

This chapter delves into the question of information order and ordering through an investigation that leads from the library, long the principal institution dedicated to the organization of knowledge, to the first widely used technique in information retrieval, coordinate indexing, and its much more elaborate cousin, the relational model for database management. The goal is not to establish a narrative of progress that inevitably leads to our contemporary situation, but to highlight certain fault lines emerging around the closely related issues of knowledge organization and access, and how they relate to specific techniques. A significant part of contemporary information ordering indeed develops from attempts to fundamentally restructure how text documents are to be cataloged and retrieved. While the library tradition sees organization as a prerequisite for access, information

Rieder, B., Engines of Order: A Mechanology of Algorithmic Techniques. Amsterdam: Amsterdam University Press, 2020
DOI 10.5117/9789462986190_CH04
retrieval rejects this idea and turns the sequence around: organization is to follow the specific needs expressed when accessing information. The order of knowledge is no longer a universal map, but a pathway defined by a momentary purpose.

The concepts and techniques involved in this story rely on different forms of physical machinery, but my account will not be another story of punched card machines, or at least not primarily. What I am interested in are the specific schemas that combine disparate material and symbolic elements into coherent systems. Specific ways of doing are clearly more compatible with specific mechanical trajectories and may rely on the availability of certain operational capabilities. But coordinate indexing finds expression in at least three material manifestations and there are good reasons to consider it more abstractly as an algorithmic technique that delineates its own technical specificity across different implementations. The arrival of the computer marks a clear incision, but, as we will see, it becomes the ‘stage’ (cf. Laurel, 1993) for a variety of techniques rather than being tied to a single operational logic or arrangement.

Following Siegert (2007), I consider these techniques as ‘media that process the observation, displacement, and differentiation of distinctions’ (p. 31) since ordering, here, usually means arranging documents in relation to each other, physically or in a catalog, in classes or as a sequence. Knowledge organization is a practice of ‘distinction-making’ where different traditions draw not only on different techniques but pursue different objectives and incorporate different ideas about the world. While my focus, in this chapter and beyond, is squarely on the lineages that form and develop in conjunction with the computer, a broader historical perspective allows me to articulate more clearly how this is a much more complex transformation than the move from ‘analog’ to ‘digital’ suggests.

The question of how words and things are to be organized is by no means a new one and since the dawn of writing as a form of material inscription, practical answers entail a tight connection between conceptual and material elements. Scholars of literacy like Jack Goody have long argued that writing itself is instrumental to the ‘process of dissection into abstract categories’ (Goody and Watt, 1963, p. 331) that dominates Western culture to this day. The earliest surviving texts, written in the fourth millennium BCE in Mesopotamia, were, indeed, mostly administrative documents in list form (Casson, 2001, p. 2), efforts to record – and thereby control – things by externalizing them onto a material surface. The list, Goody (1977) argues, is not just a memory aid, but a device that ‘enables [speech] to be inspected, manipulated and re-ordered in a variety of ways’ (p. 76) and thereby allows for the question
of how things should be arranged to become explicit. The list has no natural order, no authoritative narrative, but instead ‘encourages the ordering of the items, by number, by initial sound, by category’ (p. 81) or by other schemes, according to the specific purpose it would serve. But once the list is made, its material presence and cognitive substance, backed by the social authority of the writer, can produce effects that have far-reaching consequences for how the ‘listed’ domain is perceived and acted upon. The list facilitates both the process of ordering and the imposition of order onto the world (cf. Young, 2017).

In this chapter, I will not venture into the fundamental question of how human beings discern and distinguish the world they inhabit, how they make it cognitively and practically apprehensible through gestures like classification and hierarchization. Instead, I will trace a lineage dedicated to a highly explicit form of ordering, namely the arranging of written knowledge. My use of that term should certainly not be read as a commitment to a theory of knowledge that considers it to be something that can be fully externalized and stored in a scroll or a codex. Rather, the techniques I will discuss are themselves caught up in defining how written material can be made to serve learning and knowing, how its value can be understood and exploited. Indeed, writings of all kinds were collected and preserved from the beginning, testifying to their perceived importance and raising the question of how these potential knowledge reservoirs should be arranged.

Libraries and Knowledge Ordering

The earliest way to impose a certain order on collections of clay tables probably did not involve written lists, but simply spatial arrangement by shelf or by room. And spatial arrangement remains a central element in various forms of ordering to this day – including in the list itself, which can be seen as a specific way of arranging items on a writing surface. As far as we know, lists of written works appeared roughly around 2000 BCE, when collections began to grow beyond a certain size. These early inventories did not follow any particular order, but the mere fact that there was an easy way to consult the holdings of a library should be considered a first step toward systematization (Casson, 2001, p. 4). The roughly 30,000 tablets in Assyrian king Ashurbanipal’s library, created in the seventh century BCE and archeologically preserved due to its destruction during the fall of Nineveh in 612 BCE, were arranged by subject matter, but the catalogs seem to have served mainly as inventories and vanguards against theft (Harris, 1999, p. 19; Casson, 2001, p. 13f.).
The most remarkable step toward a more deliberate approach to cataloging happened at the Library of Alexandria in the third century BCE, where Callimachus's famous bibliographic work *Pinakes* (Tables) employed different techniques to aid the finding of individual texts, inaugurating an order of knowledge in the process. This monumental catalog of the library’s 490,000 scrolls was divided into ten subject matters that contained further subcategories. Authors were, for the first time in history, listed in alphabetical order to facilitate retrieval. The *Pinakes* has not survived, but we know that the catalog divided authors into poetry and prose, and the latter into subject domains such as philosophy, history, and medicine. Individual works could be retrieved with the help of a shelf list (Casson, 2001, p. 39ff.). Systematic organization and broad access had direct consequences for the use of the library as an institution of knowledge management and acquisition, making it ‘possible for the student to plan for himself a rational course of reading in whatever subject interested him, offering opportunities for independent learning’ (Lerner, 2009, p. 17). The arrangement in subject matters projected a specific division of knowledge domains onto the most important library of antiquity, and thus onto a large zone of influence, but the enumeration of authors following the letters of the Greek alphabet stands out even further. For Goody, this mode of arrangement is a particularly powerful feature of the combination between the list form and the phonetic alphabet:

> [T]he value of alphabetic listings is that each word is automatically assigned a specific but logically arbitrary place in the system, a space that only that item can fill. It is thus of immense value in retrieval systems dealing with masses of disordered information, such as subscriptions for the telephone or students in a class. (Goody, 1977, p. 110)

While the subject catalog requires considerable intellectual effort and can be considered a genuine work of metaphysics, the alphabetical list appears as the first ‘mechanical’ ordering device. Libraries had grouped works by ‘objective’ properties such as type of support, size, or language before, but none of these allowed for the same precision, speed, and convenience. The technique only works because the (Greek) alphabet is itself an *ordered* list of letters and those who learn to read also learn the specific sequence of letters. This arbitrary ordering – there is no specific reason that \( \alpha \) should come first – represents a resource for the mechanical sorting of any set of words, providing more granularity when moving from the first letter to the second and so forth. The alphabet assigns a place to every thing through its name and remains one of the most common ordering devices
in use – every programming language ‘knows’ the sequence of letters and PHP’s sort() function will readily apply that sequence when presented with a list of character strings.

After the partial destruction of the Library of Alexandria in 48 BCE, it took almost two millennia before any institution could match the cataloging achievements of the Ptolemies. Roman libraries were much smaller in size, less systematically organized, and in large part privately owned (Casson, 2001, p. 80ff.). Libraries in the Islamic world, such as the ones in Baghdad, Cairo, or Cordoba, had book collections counting in the hundreds of thousands, but we know little about the exact details of cataloging (Harris, 1999, p. 79ff.) used in them other than that knowledge organization followed a logic of ‘nobility first’, evaluated in religious terms (Lerner, 2009, p. 62). This reminds us that classification often involves the attribution of value and significance to items.

Further to the east, the Chinese assigned great importance to well-ordered libraries and catalogs, dividing subjects into classics, histories, philosophers, and collected works following Chinese scholarly tradition (Lerner, 2009, p. 41ff.). In the West, the monastic libraries of the Middle Ages were small, focused on theological texts, and were mainly driven by the concerns of education and preservation. Since collections were so small – the library of the Vatican held a mere 1209 volumes as late as 1455 (Lerner, 2009, p. 89) – catalogs were mostly inventories of highly valuable objects and not attempts to systematize knowledge (Svenonius, 2000, p. 29; Lerner, 2009, p. 35). As monasteries prospered and expanded their educational role, ordering volumes by subject, generally following the trivium and quadrivium, became a common practice. Mostly implemented via spatial placing in library rooms, librarians ran into logistical problems as titles were added (Harris, 1999, p. 110; Lerner 2009, p. 70). Only at the end of the Middle Ages and in particular when printing started to significantly expand the number of books in circulation did catalogs become more methodical, again mostly by registering volumes alphabetically. However, as books collections grew and covered broader fields, organizational principles multiplied (Murray, 2009, p. 205), typically following the specific ideas of a head librarian who held almost exclusive power over cataloging.

Until the late eighteenth century, catalogs were themselves books that simply listed works according to one of several available ordering schemes. While librarians would leave some space between entries for new arrivals, these would eventually fill up and require notes on the margins, the addition of new pages, supplementary volumes and, sooner or later, a full – and thus very costly – retranscription of the entire catalog. But the increase in
printed volumes in the late eighteenth century and the steady stream of innovations in mechanical printing during the nineteenth century set the stage for a new material substrate that solved many of the basic limitations of the book form and stimulated interest in more systematic organization: the card catalog.

The Card Catalog

Until the Enlightenment period, state-owned libraries in the West were assembled for the purpose of representation rather than the efficient management of knowledge. The many problems cataloging and shelving encountered mattered less to the limited audiences that had access in the first place. During the eighteenth and nineteenth centuries, however, this conception of the library came under attack from democratic reformers that requested public admission to what was increasingly perceived as a valuable resource for learning and social reform (Harris, 1999, p. 243f.). Librarians fully embraced the democratic ethos of the Enlightenment and began to see themselves as stewards of knowledge that should guide seekers through ever larger collections. The transformation of the library into an institution of public education and a ‘temple of reason’ defines its mission to this day. But making access for wider audiences a reality was not only a political but also a logistical struggle. As Mortimer Taube, one of the information retrieval pioneers of the 1940s and 1950s, argues, classification became an essential part of the solution:

The patron of the public library was the ‘common man,’ the citizen of the republic of letters to whom all knowledge was to be made available. To be sure, there have always been attempts to classify knowledge; but library classification became a major factor in bibliographic organization only when it became a specialized tool for the modern public library. (Taube, 1951, p. 64)

If the new patrons were to be served well, a uniform, systematic, and up-to-date catalog was required, not only to help with orientation and book retrieval but also with management tasks such as inventory handling or lending. The card catalog greatly facilitated all of these functions.

1 As Carla Hayden, the current Librarian of Congress, writes in her welcome message: ‘The Library preserves and provides access to a rich, diverse and enduring source of knowledge to inform, inspire and engage you in your intellectual and creative endeavors’ (Hayden, n.d.).
Although the Swiss naturalist and bibliographer Conrad Gessner had already used a complex system of paper slips to compile his *Bibliotheca Universalis* in the sixteenth century, the first real card catalog was introduced by diplomat, politician, and librarian Gottfried van Swieten in 1780 in the Hofbibliothek in Vienna (cf. Krajewski, 2011). In this quite revolutionary system, pioneered in parallel in revolutionary France, works were no longer listed in bound volumes, but each title received a card and these cards would be stored in adapted cabinets, ordered by alphabet or subject. The system greatly facilitated the creation, handling, and maintenance of catalogs, in particular for larger collections. When a new title arrived, the corresponding card could be lodged between two others; if space became scarce, the cards would simply be transferred to a larger cabinet. Leaving blank space for future additions or scribbling new titles on margins was no longer necessary. While the card catalog was not directly related to any specific ordering system, it transformed the very act of cataloguing from a special project that would take place at the initiative of a particularly ambitious librarian at a specific moment in time and then slowly deteriorate into an ongoing process that could infiltrate much more deeply into library practices. The importance of the new approach is hard to overstate.

But the card catalog is not only a practical tool that facilitates the life of both librarian and library patron; it is a critical step toward mobility and abstraction, that is, toward ‘the process of considering something independently of its associations or attributes’ (OED). Catalogs in book form had already made it possible to move beyond the crucial limitation of ordering by physical placement on shelves, namely that there could only ever be a single arrangement. Creating lists of references pointing to physical locations allowed for alphabetical, subject-based, and other ordering rationales to coexist in the context of a single collection. Catalogs are essentially collections of metadata that decontextualize individual items and encourage experimentation with different forms or order. Even if card catalogs were still mostly ordered by author or subject (Harris, 1999, p. 147), the mobility of individual references further enhanced organizational possibilities. While physical books could always be rearranged but only into a single order at a time, written list could be easily multiplied but were hard to expand or adapt. Moving to cards, the reference entries themselves became movable and what held them together was no longer a list, but a cabinet. In a sense, the card catalog still follows the list form, but the drawer replaces the page and the sides of the cabinet the outer binding of the book. The major difference is that the items on the list become ‘atoms’ that are suddenly able to move, giving a boost to the reordering function.
highlighted by Goody (1977, p. 76), even if the card catalog mostly served to uphold existing ordering schemes as the mass of items to manage grew at increasing speed and the mandate of the library had begun to transform.

Although Simondon’s work focuses on more industrial forms of technology, I would argue that a card catalog is a technical individual that combines various elements – one of them being an ordering scheme – into a working system. The specific technicity of the card and cabinet had important consequences: the card catalog played a central role in transforming the library into a knowledge resource and learning tool open, at least in theory, to an entire population. The change in materiality from the book form to the card form contributed to a shift of attention from the cumbersome logistical work of cataloging to more conceptual aspects of organization and arrangement. The logistical ‘unburdening’ kept collection growth manageable and meant that ordering principles, seen as means to make collections more accessible, could come more strongly into view. Although the adoption of card catalogs was relatively slow and required the greater part of the nineteenth century, it is no surprise that their diffusion coincides with the period where the great knowledge systematizations that structure library collections to this day took form. These attempts to create stable and all-encompassing orders of knowledge were greatly inspired by the universalist outlook of Enlightenment philosophers, even if the realities of library management meant that concessions had to be made.

Toward Universalism

Next to serving as objects of prestige, state-owned libraries before the nineteenth century were (sometimes) to assist ‘persons of merit and knowledge’ (Lerner, 2009, p. 91) in their intellectual endeavors. Hiring a scholar of renown as the librarian could further both of these objectives. For someone like the philosopher and polymath Gottfried Leibniz, who served as librarian for many years, such an assignment was an excellent opportunity to think about and experiment with the organization of knowledge. His work on the catalog in Wolfenbüttel followed in large part the 21 subject categories laid out by Conrad Gessner and his own recommendations for library classification systems favored pragmatic schemes based on university faculties (cf. Schulte-Albert, 1971). He did, however, introduce a decimal extension to distinguish between the volumes of an ongoing series (Schulte-Albert, 1971, p. 140). But Leibniz was not just interested in the systematization and accessibility of the works in the institutions he headed. The ambition to devise a more perfect form of reasoning than natural language with its
many ambiguities echoes though Leibniz's designs for an ‘alphabet of human thought’ (*alphabetum cogitationum humanarum*), a system combining a universal language (*characteristica universalis*) and a universal framework for logical calculation (*calculus ratiocinator*). These elements amount to ‘not just a theory of classification […], nor just a general theory of philosophical method […], but a formalism for the representation and for the generation of all knowledge’ (Mittelstrass, 1979, p. 604). Here, we find the most explicit expression of the idea that calculation can be a means for producing (philosophical) knowledge (Krämer, 1988, p. 100f.). The attempt to develop not merely an inventory of knowledge, but a thorough formalization of both semantic representation and logical inference opens a direct lineage toward efforts in areas such as symbolic artificial intelligence or the Semantic Web.

We will come back to such intersections between calculation and meaning, but Leibniz’s work can already illustrate the interest in the organization and formalization of knowledge philosophers began to take in the seventeenth century. This interest funnels into a second lineage next to the library that was concerned with collecting and organizing human knowledge: the Encyclopedia.

Here, Francis Bacon’s *Of the Proficience and Advancement of Learning* from 1605 is worth mentioning since its classification of all knowledge into a tree with three main branches – memory, imagination, and reason – became highly influential (Darnton, 1999). Among others, Diderot and D’Alembert’s *Encyclopédie*, a highly ambitious attempt to collect and systematize all available knowledge, drew on Bacon’s system for its own hierarchical taxonomy. While encyclopedias and libraries differ in fundamental ways, they follow a similar mission and their histories intersect frequently. The *Discours préliminaire* from 1751 defines the task Diderot and D’Alembert set out for themselves, and it is worth quoting a particularly instructive passage in full:

[The encyclopedic arrangement of our knowledge] consists of collecting knowledge into the smallest area possible and of placing the philosopher at a vantage point, so to speak, high above this vast labyrinth, whence he can perceive the principal sciences and the arts simultaneously. From there he can see at a glance the objects of their speculations and the operations which can be made on these objects; he can discern the general branches of human knowledge, the points that separate or unite them; and sometimes he can even glimpse the secrets that relate them to one another. It is a kind of world map which is to show the principal countries, their position and their mutual dependence, the road that leads directly from one to the other. (D’Alembert, 1995, p. 47)
For Diderot and D’Alembert, the tree of knowledge could be drawn in different ways, but the task of the *philosophes* was not to merely produce an arbitrary arrangement supporting the pragmatic goal of making the 70,000 articles in the *Encyclopédie* accessible. In line with Foucault’s description of the classic *épistémè*, they were confident that rigorous analysis would be able to adequately and accurately account for a world ‘offered to representation without interruption’ (Foucault, 2005, p. 224). But as Darnton (1999, p. 191ff.) argues, the specific way categories were chosen was in itself a statement designed to promote secular and antiauthoritarian ideals. The universalism of the French Enlightenment is thus first a commitment to the universal values and natural rights\(^2\) that find their most widely disseminated expression in the *Déclaration des droits de l’homme et du citoyen* of 1789. The adherence to these ideals, however, also had a ‘universalizing’ effect in epistemological terms, in the sense that those who reject the authoritarian impulses of religious or political despotism and allow themselves to be guided by their desire for truth would be able to come to a shared understanding concerning the systematization of knowledge, even if an element of arbitrariness remains. The argument the *philosophes* made was twofold: on the one side, nature itself possesses a structure, which we are able to perceive through our senses and to represent adequately; on the other side, the practical task of diffusing knowledge to everyone requires a certain coherence and unity. The *Encyclopédie*, a colossal effort that drew on the contributions of 140 individuals, should thus be seen as a socio-epistemological project that attempts to produce not only a collection of all human knowledge but also the adherence to a shared understanding concerning its systematization. Epistemological and political universalism go hand in hand in this endeavor.

**Modern Library Classification Techniques**

Although libraries had to deal with logistical problems that did not affect the creators of the *Encyclopédie*, the public library movement of the nineteenth century embraced a similar set of beliefs. The objective to cover all areas of knowledge and to provide access to everyone required enormous effort and provoked increased interest in systematization, commonality, and standardization. While the librarians of the Middle Ages and Renaissance

---

\(^2\) For the encyclopedists, it is the ‘cry of Nature’ against injustice that compels an understanding of good and bad, creating the moral imperative to use knowledge as a means for emancipation (D’Alembert, 1995, p. 12).
had begun to create union catalogs that covered more than one institution and wrote manuals for library organization that were instrumental for developing a set of shared ideas, these undertakings deepened in the nineteenth century. Anthony Panizzi’s 92 bibliographical rules for the British Museum (1841) and Charles Coffin Jewett’s 39 rules for the library at the Smithsonian Institution (1853) were not only much more detailed than earlier recommendations but were also taken up in many other locations (Lerner, 2009, p. 177). These efforts culminated in Charles Ammi Cutter’s *Rules for a Printed Dictionary Catalog* (1876), ‘likely the most cited text in the bibliographic literature’ (Svenonius, 2000, p. 16), which built its rules and recommendations on three explicitly stated objectives for catalogs: to help find a specific book; to show what a library has on offer, for example, on a specific subject; to ‘assist in the choice of a book’ (Cutter, 1876, p. 10).

Such attempts to formulate explicit rules and guidelines were part of a larger conversation that led to the emergence of library science as an academic discipline and to librarianship as a profession. A crucial question was how to best realize the shared objectives Cutter synthesized into paradigmatic form and systematization was to play an important role. Other than basic author lists, the two central lines of development for cataloging were the structured subject heading list and the hierarchical classification scheme, both of which endure until today. Building on a long line of precursors, some of them already mentioned, they received their modern form in the second half of the nineteenth century.

The formulation and standardization of modern subject headings is generally attributed (Kilgour, 1969, p. 31; Wiegand and Davis, 1994, p. 605), at least in the American context, to Jewett, still working at Brown University in the early 1840s, and later to Cutter’s (1876) canonical formulation. The modern form generally contains three elements that are combined into a single, alphabetically ordered list: broad subjects, specific subjects, and so-called subject-word entries that take the word judged most important from a book’s title. The first two are taken from a controlled, uniform vocabulary and are the most important part of the system. Cutter’s version already contained the ‘see’ (for linking nonuniform terms to the corresponding uniform terms) and ‘see also’ (for linking to related or more specific headings) type references that continue to exist, for example, in the Library of Congress Subject Headings (LCSH), prepared since 1898. LCSH entries are strictly controlled expressions containing a main heading and, possibly, a subdivision (e.g., ‘Classification--Books’). As Svenonius (2000, p. 179) remarks, the LCSH is a largely enumerative language, where all allowable expressions are given in a centrally controlled list, even if more ‘synthetic’ features were added
after 1974, allowing ‘free-floating’ subdivisions, for example, for time or geography (e.g., ‘Art--Censorship--Europe--Twentieth century’).

While the preparation of an actual catalog requires consistency and thus an authoritative thesaurus of subject headings, such thesauri are generally not based on a structured, philosophical vision concerning the order of knowledge (Wiegand and Davis, 1994, p. 606). Their specific brand of universalism is conventional rather than Hegelian. Despite continuous criticism stemming from the extreme reluctance to adapt existing vocabulary (Wiegand and Davis, 1994, p. 606; Drabinski, 2013, p. 97), the ongoing use of the LCSH in the US, the Répertoire d’autorité-matière encyclopédique et alphabétique unifié (RAMEAU) in France, or the Schlagwortnormdatei (SWD) in Germany shows that the adherence to a shared standard remains highly attractive to librarians committed to the practical unification of knowledge promoted by the *philosophes*. Subject headings also work very well with the logistics of the card catalog: subjects are written on the top of taller cards and individual titles are ranged alphabetically behind them. But they largely leave open another problem that the creators of the *Encyclopédie* did not have to deal with: the placement of titles in a library.

While systematic classifications based on hierarchical category trees can be used for cataloging and indexing, their enduring success in libraries mostly concerns the physical arrangement of books. Despite the affinity between Enlightenment thinking and the library movement, it was not the Bacon-*Encyclopédie* lineage that became dominant but a much more utilitarian system that took some inspiration from Leibniz’s decimal numbering system. As its name implies, the Dewey Decimal Classification (DDC), first introduced by Melvil Dewey in 1876, is based on a nested tree structure where ten main categories are subdivided into ten subcategories, and so forth. Although the initial version only proposed a depth of three, there is no theoretical limit. Dewey, who insisted on copyrighting his invention, argued that ‘philosophical theory and accuracy have been made to yield to practical usefulness’ and conceded that ‘the division of every subject into just nine heads is absurd’ (Dewey, 1876, p. 4). But his invention allowed for a very convenient numbering scheme: Bacon’s work would be found in the main class ‘Philosophy’ (100) and the subclasses ‘Modern Philosophies’ (190) and ‘English’ (192). Although there was, from the beginning, an alphabetical index of all classes, the system’s decimal requirement introduced all kinds of conceptual difficulties that made it less suitable for cataloguing than the ‘flat’ structure of subject headings. Its hierarchical organization was more useful for the physical placement of books, which was previously based on fixed rather than ‘logical’ locations. The decimal system provided a practical
means to organize and locate books in a way that titles treating a similar subject would be placed together on the shelf, encouraging browsing and discovery. The capacity to move from the more general to the more specific traded the greater flexibility of subject headings for easier navigation. Since both systems have their respective strengths and weaknesses, many libraries employ both in conjunction.

To complete this short overview of the backdrop against which algorithmic ordering techniques develop, the work of two other pioneers is worth mentioning: Paul Otlet and Siyali Ranganathan. Although Paul Otlet’s contributions are often mentioned as precursors to contemporary forms of information access, his work can be seen as the culmination of the nineteenth century when it comes to knowledge organization. Working with international lawyer Henri La Fontaine in Belgium, Otlet founded the discipline of ‘documentation’ as a direct response to the proliferation of publications of all kind, ranging from scientific journals to audiovisual contents. The challenge to the book as the primary support of knowledge justified their highly systematic approach. In the 1890s, they founded the Office international de bibliographie and started to work on the Répertoire bibliographique universel, a ‘universal catalog’ (Otlet, 1934, p. 241) that would cover all document types. To give structure to this immense edifice, a modified version of the DDC was used. The Classification decimale universelle (CDU), still in use today, extended Dewey’s classes to account for certain particularities of the European context and introduced a syntax for cross-referencing between classes, similar to the ‘see also’ connections used in subject headings.

But Otlet was dissatisfied with mere bibliography. While recorded knowledge obviously had to have some kind of support, books and other documents contain many different ‘facts’ that are not easily accessible through a catalog. Otlet therefore also imagined an encyclopedia, a ‘universal book’, that would be based on a repertoire of facts, extracted and retranscribed on index cards and classed into the universal systematization of the UDC (Otlet, 1934, p. 7). The Mundaneum, which housed the different card collections and was to be at the heart of a network of information centers, could indeed be described as an ‘analog search engine’ (Wright, 2008) that responded to 1500 information queries per year during its heyday. Otlet indeed combined the basic principles of the card catalog with the encyclopedic impulse inherited from the Enlightenment to produce a collection of knowledge closer to the ‘logic of the database’ than the monographic form (Hartmann, 2015, n.p.). He also shared with many of his successors the idea that books were now only one support among many and that the sheer mass of publications required
creative solutions, including technical innovations such as microfilm and electronic data transmission (cf. Hartmann, 2012).

Where Otlet remains largely attached to the nineteenth century is the strong emphasis on systematization, standardization, synthesis, unity, and 'universalism, that is, a system that would embrace all things' (Otlet, 1934, p. 430). Trained as a lawyer rather than a scientist or librarian, he espoused a deeply positivistic vision of science that, as Hartmann (2012, p. 56) remarks, tends toward an almost Hegelian totalization, where the stock of knowledge forms a coherent, unambiguous set of laws that can be captured in a singular structure. Knowledge, broken up into atoms collected on index cards, would be put together again, not by the king's horses and men, but through universal classification ‘conform to the order intelligence discovers in things’ (Otlet, 1934, p. 41). Although Otlet showed a certain pragmatism when it came to logistical solutions, he shunned the alphabetical subject-heading approach as ‘too empirical’ (Otlet, 1934, p. 381) and too scattered to capture the complex expressions of contemporary science. Ironically, Otlet was heavily criticized by one of his contemporaries, Henry E. Bliss (1935), for his decision to adopt a modified version of the DDC which, ‘disregarded, or ignored, the principles of logical classification and the systems of science and education’ (p. 100). For Bliss, the UDC was not positivistic enough: if ‘[o]rganizations of knowledge […] become organizations for thought’ (p. 87) the microcosm of mental organization needs to correlate with the ‘intrinsic physical organization of the cosmos, the macrocosm’ (p. 102).

In fact, the philosophical structure of classification was not at the center of Otlet’s interests. Rayward (1994) rightly remarks that his objectives were ultimately more ambitious and classification ‘was simply the first step in a more general system of what might be called documentary processing’ (p. 237), a practice that would go deeply into the content of a document. His ideas concerning techniques for establishing connections between excavated facts are sometimes presented as precursors to hypertext and the web (Hartmann, 2006), but despite clear similarities, I would argue that his vision more accurately amounts to a centralized and fully systematized version of Wikipedia, a universally accessible, navigable repository of all ‘factual’ knowledge.

In contrast, a system that focused explicitly on pushing the mechanisms of classification beyond their nineteenth-century formulations was introduced in 1933, a year before Otlet published his magnum opus, the *Traité de la documentation*. Invented by S. R. Ranganathan, a mathematician who somewhat accidentally became librarian at the University of Madras in 1923, the 'colon classification', named for the prominent use of the colon sign in its
syntax, was ‘designed to overcome the difficulty enumerative languages have in keeping pace with knowledge’ (Svenonius, 2000, p. 175). Enumerative languages, like the DDC or the LCSH, are based on the principle that all classes have to be listed in a main thesaurus; synthetic classifications – of which the colon classification was the first fully developed example – instead define a syntax for combining ‘aspects’ into a classificatory expression, ‘reducing the vast universe of knowledge to a set of atomic concepts and certain basic relations among them’ (Svenonius, 2000, p. 175). This meant that indexers did not have to wait for a central authority, like the Library of Congress, to update their category listings. Star argues that the colon classification thus frames ‘the universe(s) of knowledge as potentially infinite, open, and evolving’ (Star, 1998, p. 226), even if the basic ordering into disciplines does not stray very far from the Baconian tradition. Ranganathan’s system is based on a list of main classes (e.g., B Mathematics, R Philosophy) that are complemented by five ‘facets’, each marked by a specific sign: ‘;’ for personality, ‘;’ for matter, ‘:’ for energy, ‘.’ for space, and ‘”’ for time. While both the DDC and the CDU have known facets in some sense from their inception, for example, to distinguish between types of publication or to indicate period and geography, the colon classification fully embraces the synthetic principle (Satija, 2002). Facets are independent from each other and can be combined into compound classificatory expressions. The technique does not, however, define a singular abstracted system of semantic coordinates. For example, personality in the context of library science means the kind of library, while in the context of chemistry it describes elements or compounds. But the different components can form complex phrases, such as the commonly given example ‘research in the cure of the tuberculosis of lungs by x-ray conducted in India in 1950s’ (Chan, 1994, p. 39), which is noted as L₄₅;421:6;253:f.44’N₅ and translates, from left to right, into ‘Medicine’ (main category), ‘;Lungs’ (personality, the specific subject), ‘;Tuberculosis’ (matter, the property), ‘;Treatment’ (energy, the intend to cure the tuberculosis), ‘;X-ray’ (matter, the means to cure), ‘;research’ (energy, this concerns research into this specific treatment), ‘;India’ (space), and ‘;1950’ (time). This system allows for extremely fine-grained descriptions and makes it easy to add new elements to different facets, for example, a new treatment. In certain respects, Ranganathan’s system is closer to Leibniz’s characteristica universalis than to traditional classification systems. While the colon classification has remained in active use, in particular in the Indian context, its enormous complexity makes it hard to use beyond special cases. It does, however, point into the broad direction algorithmic information ordering was taking.
Toward Computation: Documentation and Coordinate Indexing

At the eve of WWII, two techniques for organizing documents by subject dominated: subject headings, generally based on increasingly long lists of increasingly specific topics (e.g., ‘Classification--Dewey decimal’ in the LCSH), and classification trees, often used for physical placement, where each book is supposed to sit at a specific place in the hierarchy. This is the backdrop against which the field of information retrieval developed a set of ideas, concepts, and – most importantly – techniques that broke with the library tradition in fundamental ways. Often relying on statistical approaches, information retrieval became one of the fundamental ‘ancestral communities of probabilization’ (Mackenzie, 2017a, p. 118) and its contributions continue to define contemporary information ordering. The rest of this book can thus be understood as an attempt to ‘excavate the present’ (Parikka, 2011, p. 214) by following a series of inventions that sought to transform knowledge organization into a dynamic, machine-supported process of ordering.

Documentation Centers

Even though catalogs, classification schemes, and other aspects of library logistics had become quite elaborate in the first half of the twentieth century, the degree of mechanization remained low. Herman Hollerith had begun to commercialize his punched card tabulators in the 1890s, but as Williams (2002, p. 16) notes, the earliest reports of uses in libraries came in the mid-1930s and concerned management tasks such as circulation control and the statistical analysis of book use to inform future buying decisions. Existing catalogs relied heavily on trained librarians as ‘knowledge mediators’, both for bibliographic control when new titles arrived and to guide visitor through standardized vocabularies and intricately organized floor plans. Even university libraries, catering to more specialized audiences than their public cousins, were organized in similar ways and run by professional librarians rather than academics. Their broad outlook and public mandate meant that the mechanization of information ordering did not happen in these institutions, and neither did the upheaval of conventional ordering schemes.

The emergence of technical and intellectual innovations – or disruptions – in knowledge ordering are closely linked to so called ‘special libraries’ and ‘documentation centers’ (Williams, 2002; Varlejs, 2004). Special libraries

---

3 Some libraries had also begun to use alphabetical lists mixing authors, titles, and subjects into a single catalog.
had existed at least since the early 1800s and they had generally been more concerned with matters of classification than their less specialized cousins (Shera, 1952, p. 189f.). But the emphasis on public education had marginalized these more focused institutions, with ‘the disastrous effect of diverting librarianship from its proper concern with the analysis and organization of recorded knowledge’ (Shera, 1952, p. 191). In the US, documentation services had not received the same amount of attention as in Europe, where Paul Otlet was promoting the idea of information centers, and were mostly concerned with new methods for photographic reproduction (Shera, 1952, p. 193). However, these traditions ‘received a new vitality when the Second World War brought into existence a need for greater and more efficient access to information than traditional library methods were able to give’ (Shera, 1952, p. 195). Varlejs gives a concise summary of the background to the substantial change in perspective and approach about to happen:

Hundreds of thousands of reports generated by wartime and postwar scientific and technical research, together with several hundred tons of captured enemy documents, challenged traditional methods of information organization, dissemination, and retrieval. In the interests of maintaining military superiority and jump-starting the economy, U.S. government policy favored maximum declassification and speedy dissemination of information to business and industry. The urgent need to carry out this policy fostered new techniques and adaptations of old methods in the 1940s and 1950s, some of which have had a lasting impact. (Varlejs, 2004, p. 89)

Existing as well as newly founded centers quickly became the practical and intellectual locus of a movement of ‘modernization’ that would concern all aspects of handling documents and, by extension, of handling information and knowledge. While the tasks at hand were still often referred to as ‘library problems’ (Mooers, 1950, p. 2; Maron, 1963, p. 144; Rau, 2007, p. 159), their main object was not so much the traditional book, but rather technical and scientific literature, generally published in the form of research papers or reports; the real and imagined audience was not the ‘common man’, but scientists and engineers in academia, government, and business; and the political mandate was not public education, but scientific, economic, and military development in the context of the Cold War, which affected the development of computing more broadly (cf. Edwards, 1996).

4 In his foreword to the Weinberg Report, US president Johnson writes: ‘One of the major opportunities for enhancing the effectiveness of our national scientific and technical effort and
The already mentioned idea that there is a ‘problem of too much information’ (Bowles, 1999, p. 156), a ‘growing mountain of research’ (Bush, 1945, p. 112) or simply a ‘scientific information crisis’ (Weinberg, 1963, p. 15) is virtually everywhere in the hundreds of papers and reports published in the wake of WWII that begin to flesh out a technical and conceptual domain that was initially referred to as ‘machine documentation’ (Wasserman, 1965, p. 19) but ultimately consolidated around the productively vague concept of ‘information’. The idea that too much information is detrimental to efficiency and progress became the founding myth of what Calvin Mooers (1950), one of the central figures of the emerging field, canonically named ‘information retrieval’ in 1950. If the operational goals of libraries since the eighteenth century were closely tied to the ideals of the Enlightenment, the explicit objective of special libraries and documentation centers was to make knowledge productive and ‘actionable’ in a direct and measurable sense and as quickly as possible. In a climate of technological optimism, this inevitably meant that machines were to play a central role.

The centers and institutions dedicated to what a much-regarded presidential report in 1963 called the ‘transfer of information’ (Weinberg, 1963) often collaborated with academics and with startups that developed in close physical, social, and intellectual proximity to universities and other research bodies. The initial conviction that solving the information crisis would be beneficial to development in just about all sectors of society meant that considerable amounts of money were made available through government agencies. And these funds were often distributed in the form of grants promoting experimentation and fundamental research. While the effects of the concerted governmental support on the emergence of computing post-WWII is generally well-known (cf. Campbell-Kelly and Aspray, 1996), backing from direct funders, such as the Office of Naval Research, and intragovernmental institutions for the promotion and coordination of research, such as the Research and Development Board, also benefited the

the efficiency of Government management of research and development lies in the improvement of our ability to communicate information about current research efforts and the results of past efforts’ (Weinberg, 1963, p. iii).

5 This shift is completed on an institutional level when the American Documentation Institute, founded in 1937, changed its name to American Society for Information Science in 1968. In 2000, it finally became the American Society for Information Science & Technology.

6 ‘The true concern of the special library is the information that its clientele will need today and tomorrow. If the special library is doing its work effectively, its clients’ needs will change rapidly, in response to the ever-changing environment that it has helped to create’ (Lerner, 2009, p. 156).
wider field of documentation and information retrieval, whether it aimed directly at mechanization or not.

The role of Vannevar Bush merits specific mention in this context. Bush, a pioneer in analog computing, became a prominent science administrator during the war, heading the Office of Scientific Research and Development, which oversaw the Manhattan Project. Both his research in computing and his work in science management prompted an interest in the use of machines to support the organization of and access to knowledge, with the goal of facilitating the generation of new knowledge. In an often-cited article from 1945, he described a machine, the Memex, that would allow its users to retrieve and connect documents, stored on microfilm, via ‘associative trails’, networks of connections that evoke the subsequent notion of hypertext. While the Memex, a personal information device rather than a retrieval system, probably had little actual influence on concrete research in the field (cf. Buckland, 1992), Bush's considerable clout, as both the 'engineer of the American century' (Zachary, 1999) and influential figure in research funding, contributed to securing public interest, recognition, and financial support for everything having to do with mechanical information processing.

Taken together, these elements led to an environment where special libraries and documentation centers with higher budgets and smaller, more specialized audiences than their publicly minded counterparts were in a position to experiment with a variety of technical innovations (Williams, 2002, p. 16). These institutional differences were accompanied by a fundamental divergence in outlook that led to what Bowles (1999) termed the 'information wars', a rift that can be traced up to the present. To understand this divergence within a broader historical frame, we have to examine the role calculation and mechanical reasoning have played in business and government long before digital computers were invented.

**Mechanical Reasoning, Statistics, and Decision-Making**

In the introduction to their seminal *Computer: History of an Information Machine*, Campbell-Kelly and Aspray argue that the ‘canonical’ story of the invention of the computer as a calculating device born out of the encounter between centuries of mathematical innovation and America’s WWII efforts is incomplete. They point out that ‘research scientists and atomic weapons designers still use computers extensively, but the vast majority of computers are employed for other purposes, such as word processing and keeping business records’ (Campbell-Kelly and Aspray, 1996, p. 2) and propose to incorporate the mechanization of office work and data processing into
the history of computing, for example, the famous application of Hollerith machines to the 1890 US census.

Although there are clear connections between information retrieval and tasks like data processing and document management, there are even broader traditions of applied mechanical reasoning situated at the intersection of calculation and decision-making that are largely absent from the (pre)history of computing. Editors Becker and Hayes (1963) indeed introduce their influential Information Sciences Series, published by Wiley & Sons from 1963 onwards, with the statement that ‘[i]nformation is the essential ingredient in decision making’ (p. v), connecting the newly established ‘resource’ directly to management practices in business and government.

Decision-making assisted by calculation has a long history, and the development of capitalism, in particular, is closely related to the use of arithmetic as a means to take on size, complexity, and uncertainty. The fifteenth century is particularly rich in this respect: Luca Pacioli’s *Summa de arithmetica, geometria, proportioni et proportionalità* standardized and disseminated double-entry bookkeeping while popular *algorismi*, manuals for learning arithmetic, proposed practical methods that were both enabling and responding to the needs of increasingly complex forms of trade, such as dealing with logistics and planning, with diverse units and currencies, and with the distribution of risks and profits. The requirements of long-distance trade, the emergence of larger commercial entities, and a general rise in organizational complexity ‘elevated computation to the status of an empirical science’ (Swetz, 1987, p. 295). This science provided precise tools to control and to decide, which stabilized, standardized, and systematized how merchants managed their businesses and interacted with their peers. While these methods were not (yet) the scientific management that emerges in the late 1800s, the considerable risks that characterized early capitalism inspired new forms of control relying on data collection and calculation (cf. Beniger, 1986).

These techniques – listing and tabulating, applied arithmetic with Arabic numerals, bookkeeping – subsequently spread to other domains. William Petty’s *Political Arithmetik*, developed over the second half of the seventeenth century, is of particular interest as it systematically applied quantitative reasoning to matters of government and produced concrete recommendations concerning public investment and economic policy. Rather than (merely) appealing to moral principles, to the authority of great individuals, or to ‘evident’ truths, Petty’s mode of argumentation amounted to what we would nowadays call cost-benefit analysis. As the following
problem from one of Petty’s essays (1655, p. 63ff.) shows, the discussed topics ring eerily familiar, even if the details of the calculation certainly seem peculiar: given the worth of a human being, based on the price of a slave at the market in Algiers, would additional spending on certain Parisian hospitals to lower mortality rates represent a good investment for the king of France? Answer: yes.

It is symptomatic of our somewhat lopsided historical perspective, which puts abstract and seemingly disinterested pursuits over practical and profitable ones, that Charles Babbage is now mostly remembered for his work on mechanical calculation, although his considerable contributions to statistics, economics, insurance, and management – the foundations of the considerable fame he won during his lifetime – could be considered as equally important to the history of computing. After all, one of the main uses of his difference engine would have been the calculation of life expectancy tables for the purpose of setting the price of life insurance annuities (cf. Babbage, 1864). Certainly, these fields were not directly related to the mathematical theory of computation and its mechanization in modern computers, but they provided a space beyond science and engineering where such machines could find an objectal embedding through some form of practical or imaginary utility. This tension between the fundamental and the applied is characteristic and the history of statistics is of particular relevance in this context, since it arguably represents the most pervasive application of calculation to governance and decision-making.

To this day, the term ‘statistics’ has kept a double meaning. First, it refers to the collection of facts or data, which, in the nineteenth century, turned into a true ‘avalanche’ (Hacking, 1990, p. 2) or ‘deluge’ (Cohen, 2005, p. 113) of tabulated numbers covering every conceivable subject. This kind of description by numbers is what we mean when we talk about ‘accident statistics’ or ‘employment statistics’ and it directly connects to various technical inventions and organizational practices emerging after 1800 (cf. Gardey, 2008). Second, ‘[b]y 1889, users of statistics [...] had become discontent with the mere presentation of numerical facts, and were looking for more refined methods of analysis’ (Gigerenzer et al., 1989, p. 58), which led to the development of statistics as a series of mathematical concepts and techniques used to find and analyze patterns in collected data, for example, dependencies between variables. Both meanings refer to ‘epistemic practices’, that is, practices caught up in the production and definition of knowledge, but the second indeed mobilizes mechanical reasoning in its fullest sense, as a purely formal transformation of symbols that nonetheless produces an epistemic surplus. The detection of a significant level of correlation
between two variables is not simply a ‘presentation of numerical fact’; it must be considered a ‘cognitive’ operation that generates an interpretation of the world behind the numbers, an interpretation that consistently lays claim to being both true and enlightening, and thus to a legitimate role in orienting ideas and actions. In the words of economist and engineer Jules Dupuit (1995), written in 1844, ‘mathematics [...] are machines that, in a certain sense, can think for us’ (p. 92). Whether we fully accept this claim is secondary here; what counts is that ‘in the operations of government, the conduct of business and finance, the activities of science and engineering, and even in some aspects of daily life’ (Cohen, 2005, p. 17), mechanical reasoning involving complex calculations has become an accepted form of relating to a world seen as complex, uncertain, and wrought with struggles for power. As Porter argues, ‘quantification is a social technology’ (Porter, 1995, p. 48), a ‘technology of trust’ (Porter, 1995, p. 15), that purports to reduce subjectivity, individual discretion, and, ultimately, the arbitrariness of the powerful, by instating ‘rules of discourse so constraining that the desires and biases of individuals are screened out’ (Porter, 1995, p. 74). Even if associated techniques and practices have come to know widespread application and acceptance, claims to objectivity and impartiality have certainly been contested. Tensions manifest particularly clearly in contemporary critiques of ‘big data’, where ‘trust in the objectivity of quantified methods as well as in the independence and integrity of institutions deploying these methods’ (p. 204) is embattled from different sides. Such manifestations of a much larger ‘legitimation crisis’ (Habermas, 1975) can be observed in many domains and ‘library problems’ are certainly one area where the question of how decisions are made and who makes them have been at the heart of arduous debates.

Independently from contestations of legitimacy, statistics constitutes the most pervasive example of machines ‘thinking for us’ that is not based on forms of logical inference, like Leibniz’s calculus ratiocinator, but on the collection and analysis of (quantitative) empirical data. The idea that diligent counting and mathematical examination of what is counted allow for a particularly performant coupling of knowing and managing is the point where statistics intersects with Foucault’s work on power. In the context of ‘bio-politics’, a ‘technology of power’ that concerns itself with the management of life and thus with birth and mortality rates, public hygiene, or urban living conditions, Foucault sees statistics as a ‘technique

7 The term is used here in connection with its etymological root in the Latin cognoscere, ‘to know’, and should not be read as a commitment to any particular theory of (human) cognition.
of observation’ (Foucault, 2007, p. 161) that plays a central role in the regulation of a population. The objective of bio-politics is not merely to keep the population in a state of submissive docility, but rather to optimize output and to ‘turn society into a machine of production’ (Foucault, 2007, p. 194). Foucault frames statistics in line with the etymological origins of the term as ‘the state’s knowledge of the state’ (Foucault, 2009, p. 162) and puts it at the center of a ‘governmental rationality’ (Foucault, 2008, p. 56) that is no longer an art of government, but a ‘political science’ (Foucault, 2009, p. 106) intrinsically tied to political economy as its dominant way of thinking society. While these associations point beyond the subject matter treated in this book, they should remind us that information ordering techniques under conditions of pervasive computerization have come to serve similar governmental functions.

The application of statistics to decision-making provided a model when it came to framing the library problem and, by extension, the problem of knowledge organization and access as a technical problem that could be solved with the proper scientific attitude and methodology. WWII certainly played a crucial role in more than one respect, but it crucially ‘strengthened the decision theoretic approach to statistical problems, and accustomed people to the idea that reasonable decisions can be made on the basis of formal, mechanized reasoning combined with measurements’ (Gigerenzer et al., 1989, p. 120). Operations research, which developed a largely probabilistic approach to military decisions during the war, left deep impressions on the individuals and institutions that set out to transform the handling of published material into a science of information. This included a broad ethos carrying ‘a novel blend of pessimism about the scope and quality of human reason and optimism about the power of social and technical mechanisms for producing rational choices’ (Heyck, 2012, p. 100). As we shall see, the mission to remove or at least rearrange ‘human reason’ in information ordering did not sway toward the universalist classifications of the nineteenth century. Espousing a probabilistic outlook that recognized uncertainty as a fundamental fact of life, it turned to statistics as a means to ‘tame chance’ (Hacking, 1990) in a way that combines the imperatives of rationalism and pragmatism.

Desrosières addresses what I mean by ‘pragmatism’ when he discusses how statistics allows for different attitudes toward the world it means to describe, ‘metrological realism’, the correspondence or equivalence theory of truth,

---

8 Operations research itself connects to a long prehistory in scientific management. See, for example, Chandler (1977) or Yates (1989).
being just one of several ‘orchestrations of reality’ (Desrosières, 2001, p. 346). What we find in commerce and government is in fact ‘accounting realism’, where the “equivalence space” is composed not of physical quantities (space and time), but of a general equivalent: money’ (Desrosières, 2001, p. 342). Here, the benchmark for validity is no longer disinterested correspondence between reality and representation but effectiveness when it comes to attaining specific goals. The notion of accounting realism thus echoes Lyotard’s (1984) assessment that the contemporary knowledge regime seeks ‘no longer truth, but performativity – that is, the best possible input/output equation’ (p. 46). Subscribing neither to the ‘high rationalism’ epitomized by Chomsky’s computationalism, nor to the epistemological and political universalism that characterizes the public library, information retrieval evolves in an intellectual environment and institutional context that espouses accounting realism as its dominant epistemological attitude.

Information Wars

Coming back to the main thread of the story, we can indeed appreciate how the early literature on information retrieval frames the ‘library problem’ as the question of how to make recorded knowledge useful. At least in the 1940s and 1950s, this debate still occurs in large parts around institutions, publications, and conferences that remain connected to a library science tradition marked by Enlightenment values such as public education and philosophical universalism. But what is underway is a full-frontal assault on this tradition. In the paper where Mooers first introduces the term ‘information retrieval’, he makes a programmatic declaration that could hardly be any clearer:

In order to approach this slippery problem with any hope of success or efficiency of thought, it will be necessary for us to put aside almost all the ideas, doctrines, and symbolic or metaphysical superstructure about libraries and library methods that we have learned or otherwise picked up in the past. It can be said – and demonstrated – that almost everything that librarians hold dear in classification is absolutely wrong for information retrieval. (Mooers, 1950, p. 4)

While these are not the first shots fired in the ‘information wars’, ‘the professional battle between scientists (documentalists) and humanists (librarians) over information retrieval’ (Bowles, 1999, p. 156), Mooers himself epitomizes certain fault lines characterizing the dispute. Trained in mathematics and
physics rather than library science, he considered the library methods of the
time to be positively archaic, ‘stalled for two millennia’ (Mooers, 1950, p. 17),
and ‘incapable of information retrieval’ (Mooers, 1950, p. 4). He saw a need
for a completely different approach ‘guided by the principles of engineering
and the scientific method instead of the outworn metaphysics’ (Mooers, 1950,
p. 17). To Mooers and his colleagues, the term ‘scientific method’ implied
not just mathematization and measurement but included the commitment
to ideas related to scientific management and, in particular, to operations
research and applications of analytical methods to process optimization
and decision-making (cf. Rau, 2013).

Even if Mooers was particularly vocal, the dismissive attitude toward
traditional library ideas was widespread. Together with the ‘information
overload’ hypothesis, the idea that there was precious little to learn from
cataloging principles and techniques like the DCC or the LCSH became the
point of departure for the emergent discipline. Mortimer Taube, another
central figure in the field, indeed argued that ‘rebels’ like himself ‘invaded
the holy of holies of sanctified library practice’ with radical new ideas (Taube,
1952, p. 167). In an article published in 1982 in the influential Advances in
Computing series, James E. Rush looks back at these early days with little
modesty:

[T]he automation revolution has resulted in much greater technological
innovation in libraries during the past 20 years than had been achieved
during the entire history of librarianship up to about 1960. (Rush, 1982,
p. 334)

Contrary to most libraries, the people staffing documentation centers were
indeed mostly scientists, not librarians (Williams, 2002, p. 20). And the ten-
sions between the two ‘sides’ were palpable.9 In his book The Librarian and
the Machine, published in 1965, Paul Wasserman euphemistically remarked
that ‘there seemed now to be a number of members of the academic com-


9 ‘Today we seem to be on the brink of startling new developments in the field of high-speed
manipulation of large masses of bibliographic data, and already the emotional responses are
becoming apparent’ (Shera, 1952, p. 15).
systems, such as the DCC.\textsuperscript{10} The ideas of non-American thinkers like Otlet or Ranganathan were discussed even less.

Similar to the statisticians of the nineteenth and twentieth century and subsequent ‘method experts’, the ‘information scientists’ of the 1950s and 1960s who pioneered many of the algorithmic ordering techniques at work today presented themselves as modernizers and rationalizers. Their story fits neatly into the picture drawn in Ensmenger’s (2010) \textit{The Computer Boys Take Over}, which details the intricate connections between the nascent and expansive discipline of computer science on the one side and the ideological and methodological nexus forming around business administration, operations research, systems analysis, cybernetics, efficiency consulting, and associated techniques. The task information retrieval set for itself was not merely to introduce new machinery and techniques into an existing setting but to redesign the library on the whole as an information system.\textsuperscript{11} And while much of the concrete work revolved around automation and mechanization, the systems perspective always lingered in the background, becoming particularly noticeable whenever new performance measurements sought to quantify the effectiveness of provided services.

These elements for a social history of the information wars can provide crucial context, but I am specifically interested in the actual techniques transforming knowledge organization into information ordering. The information wars can indeed be narrated in terms of machinery (computers vs. manual methods) or socialization (scientists vs. humanists), but the actual ‘epistemological operators’ (Young, 2017, p. 45) put forward crystallize around a specific document retrieval technique that lays the foundations for many further developments: coordinate indexing.

**Coordinate Indexing**

As already mentioned, the objective of special libraries and documentation centers was not to support public education but to facilitate access

\textsuperscript{10} This is what Mooers has to say about Dewey’s system: ‘The postulates of the Dewey system are incompatible among themselves, and the system can never be readjusted so as to perform the task set for it’ (Mooers 1950, p. 8).

\textsuperscript{11} A 1965 survey on the state of automation in libraries, porting the ominous subtitle \textit{A Study in Reluctant Leadership}, criticizes the lack of systems thinking: ‘The preceding survey shows that none of the libraries which have adopted mechanization or automation to any extent has attempted seriously to consider a whole library as a total information system which is to be integrated so far as possible in its concepts, design, provision of equipment, and daily operations’ (Gull, 1965, p. 5).
to relevant information for the ‘knowledge worker’: the research scientist pushing for new findings and the engineer requiring a particular detail about a specific process, but also the manager having to make a quick decision and the public administrator planning the details of a specific policy. With Foucault, we should hesitate to see this shift as another step in a general movement of rationalization. As we have seen, traditional library methods were certainly not immune to rationalist universalism and if we ‘analyze specific rationalities rather than forever invoke the progress of rationalization in general’ (Foucault, 1982, p. 779f.), we realize that this is indeed a matter of competing claims to rationality.

The most fundamental difference, in fact, concerns the framing of the library problem. The information specialists were not concerned with lofty ideals of emancipation through knowledge but saw themselves as contributors to professional environments where information was a central resource. Information retrieval’s purpose was thus to satisfy specific ‘information needs’, a concept that has remained at the heart of information retrieval ever since. These needs were seen as not being met by existing methods and techniques:

What is required is the recognition that the Library of Congress system, for all its complexity and detail, is not a tool for specialists but a general system for the non-specialist’s approach to knowledge as a whole. (Taube, 1951, p. 63)

Specialists required specialized means to access specialized knowledge and coordinate indexing was one of the earliest techniques built on this very premise. It developed, at least indirectly, from earlier applications of document processing techniques by scientists using edge-notched punched cards to analyze categorical research data in the 1930s. The McBee Keysort system was probably the most used commercial variant. These were ‘almost entirely “term on item” systems’ (Kilgour, 1997, p. 340) where a card would register various characteristics (terms) of a chemical compound, animal species, or plant (items). A notched card could encode as many binary variables as the

---

12 The definition of information retrieval given by one of the most well-known textbooks reads: ‘Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfy an information need from within large collections (usually stored on computers)’ (Manning et al., 2008, p. 1).

13 This statement requires some qualification since special services like abstracting or ‘current awareness’ publications can be seen as fulfilling retrieval functions and such services had been around since the nineteenth century.
space on the cards’ edges permitted. The card would ‘describe’ a particular entity by marking the value for each field with either a hole or a notch.

A card for ‘dog’, for example, would have a hole in the ‘carnivore’ field and a notch for ‘retractable claws’. The card for ‘cat’ would have a hole for both. When using a so-called sorting needle on a stack of cards along a particular characteristic, the cards with holes would get lifted while the cards with notches would stay put. If we were to apply the needle to the ‘carnivore’ field, both dog and cat would be pulled, but the needle would not get the dog card when fishing for retractable claws. While actual implementations of this
system could be entirely manual, the simple logic based on items (objects) and terms (properties) was well-suited to mechanization. And despite its simplicity, the system allowed for Boolean operators in queries: using two needles would constitute an OR; using a needle sequentially to first get items with characteristic A and then lift cards with characteristic B from those results would constitute an AND. The outcome would be a *selection*, following no particular order, of documents satisfying the specified logical condition. Even the NOT operator could be applied by removing cards with certain properties from the results (Kilgour, 1997, p. 341). Although the cats and dogs example may seem trivial, such a system was able to query thousands of items along up to around a hundred binary variables, the physical limit of the cards, which meant that it was capable of ‘operationalizing distinctions’ (Siegert, 2013, p. 14) dynamically in relatively large and complex datasets. This simple principle constitutes the operational schema in what was to become the first and, for a considerable time, the dominant framework for the computerization of document retrieval.

The technique was applied to bibliographic indexing in several places at roughly the same time, most prominently by Mortimer Taube, even if his approach favored a somewhat different implementation than the one just described. Taube’s system, for which he invented the term ‘coordinate indexing’, solved one of the problems related to a major limitation of notched cards, which could deal with fairly large collections of items but were only able to accommodate a relatively small number of terms, limiting the size of the indexing vocabulary. Taube thus decided to use cards not only for listing the documents to be searched but also for what he called ‘uniterms’, simple keywords describing catalogued works (Wasserman, 1965, p. 57f.; Kilgour, 1997). While the procedure for selecting and attributing uniterms to documents was a matter of debate, the general idea was to eschew complex categories or subject headings in favor of single words (Doyle, 1975, p. 174). Every uniterm received a card which listed the reference numbers for the works tagged with that term. A card could, in principle, list hundreds of titles and a title described by ten terms would therefore appear on ten cards. To make queries, searchers would combine or ‘coordinate’ the terms they were looking for and then search for the numbers appearing on all of the cards. Partial matches could be taken into account and counting the number of query terms a document would match constituted a simply ranking mechanism.

In the beginning, the system relied on manual use, facilitated only by the rather ingenious use of ordered columns for reference numbers to facilitate lookup. In this setting, the numbers of terms and items were mainly limited
by the searcher’s capacity (or willingness) to drudge through lists and lists of entries. What would appear completely impractical for the large stocks of public libraries, which covered many different areas of knowledge, was much more workable in the context of specialized collections. Coordinate indexing became a widespread and popular technique (Wasserman, 1965, p. 62; Salton, 1987, p. 376) and the introduction of Boolean operators continues to resonate through electronic catalogs and other search systems. Compared to traditional classifications or subject headings, coordinate indexing had two perceived clusters of advantages that each connect to larger debates about knowledge organization and access.

First, compared to existing indexing practices, the attribution of uniterms or descriptors was seen as easier, faster, cheaper and more adapted to the needs of searchers. The mandate to serve the professional knowledge worker meant that special libraries and documentation centers were not limited to books but included the quickly growing mass of journal publications and reports. Together with the desire for fast distribution of the latest findings, often termed ‘current awareness’ (Svenonius, 2000, p. 28), and thus for fast bibliographic attribution, this meant that the person tasked with indexing would have little time to spend on individual items. Much of the early literature on coordinate indexing thus builds its criticism of library techniques on the central role complex classification and subject heading systems confer to those who develop these controlled knowledge structures and those who assign records to specific places in the system. The ‘bibliographic power’ (Wilson, 1968) these knowledge mediators hold is not seen as a guarantor of neutrality but as a source of subjectivity, arbitrariness, and conservativism, which is particularly problematic in emerging and cutting-edge fields of knowledge. Since standardization and continuity were considered essential by librarians, the structures and vocabularies of systems like the DDC or LCSH were somewhat extended but never substantially reorganized. While coordinate indexing did not necessarily do away with the idea of a predefined or controlled vocabulary, since institutions often developed their own thesaurus, it opened the door for thinking about ways to eliminate the expert indexer trained in the arcane details of the indexing system.

Explicitly framed as a technique for creating any number of bibliographic systems, coordinate indexing never had any universalist pretentions. From the beginning, it embraced local perspectivism not only in the sense

---

14 ‘I think that in this matter librarians have been misled by the universal acceptance of the elements of descriptive bibliography and have concluded erroneously that it is possible to create a universal system of subjects which will be satisfactory to all specialists’ (Taube, 1951, p. 61).
that uniterm vocabularies were only meant to cover specific fields but also by empowering the expert searcher. This was accomplished by moving from static and complex ‘precoordinated’ systems to a ‘postcoordinated’ perspective, where compound, multipart concepts were no longer attributed by a librarian at the moment of indexing, but constructed by the searcher at ‘retrieval time’ through the combination of simple terms using Boolean operators.\textsuperscript{15} The complex precoordinated entry ‘Spain--History--Civil War, 1936-1939--Literature and the war’ from the LCSH would become an unordered list of keywords such as ‘Spain, Civil War, Literature’ in the postcoordinated logic. Luhn summarizes the difference in particularly clear language:

The introduction of discriminating terms, such as key words, has brought about a new degree of freedom. Instead of applying a static set of categories or subject headings beforehand, such categories or subject headings may be formulated dynamically at the instant of inquiry by a tailor-made assembly of such key words. (Luhn, 1968a, p. 126)

Keywords or uniterms could be attributed much more easily, disregarding order and syntax. The already mentioned Weinberg report (1963), for example, stressed the role of scientists in the process, asking them to ‘self-index’ their papers through relevant keywords. Providing well-written abstracts should help indexers with ‘emergent’ coding where keywords would be taken from the document itself. This, indeed, became the recommended method (Doyle, 1975, p. 176) and anyone who has ever submitted a paper to an academic journal should be familiar with these practices. Moving away from controlled vocabularies meant that librarians were not only besieged on the side of indexing but also on the side of searching, since users could employ terms they were familiar with rather than a predefined syntax and semantics (Lerner, 2009, p. 190). Overall, the measured efficiency of coordinate indexing was similar to more elaborate methods, while being faster and cheaper (Salton, 1987, p. 3).

Second, coordinate indexing was seen as ideally suited for mechanization and computerization. Since Alonzo Church and Alan Turing had shown ‘that numbers were an inessential aspect of computation – they were just one way

\textsuperscript{15} ‘Pre and post reference the time at which terms are concatenated into large expressions. In a precoordinate language, this is done prior to retrieval by a professional in a manner defined by the syntax rules of the language. In a postcoordinate language, it is done at the time of retrieval by a user using a Boolean-based syntax’ (Svenonius, 2000, p. 178).
of interpreting the internal states of the machine’ (Barr and Feigenbaum, 1981, p. 4), the ‘digital handling of non-numerical information’ (Mooers, 1950) had become a plausible endeavor. While early computers were much too costly for practical use, Wasserman (1965) argues already in 1965 that ‘an almost morbid fascination with the latest form of technology’ (p. 9) had contributed to a climate where the computerization of library functions started to look like both an imperative and an inevitability. Coordinate indexing played a central role in this context, because ‘[i]t was the only system for retrieving bibliographic information that was compatible with computers of the mid-1950s’ (Segesta and Reid-Green, 2002, p. 28). The formulation ‘compatible with’ highlights the fact that coordinate indexing is not bound to a particular material substrate. It plays the role of a technical element that needs to be articulated in conjunction with other elements to become a working system. While coordinate indexing represents a departure from the library tradition, its success relied on the straightforward compatibility with existing technical trajectories, including manual list comparison, punched cards, film, and early computing hardware. This compatibility can be attributed to its practical and conceptual simplicity, especially compared to Ranganathan’s similarly flexible but much more complicated colon classification. In the early 1950s, information processing machinery was both an incipient reality and a future promise, but Taube and others fully invested this emerging technological space, arguing that ‘bibliographic coordination is designed for machine sorting and collating’ (Taube, 1951, p. 70), even if it was not clear which technical trajectory would eventually dominate.

The limited space for terms on edge-notched cards prompted Mortimer Taube to develop his list-based approach, which allowed for the creation of what we would today call ‘high-dimensional datasets’, where each item can have a large number of variables or properties. The very same limitation led Mooers to develop a system that was relying on similar principles but used a different implementation. At a time when Taube, a trained librarian, was promoting his manual technique using concepts and terms familiar to librarians, the mathematician Mooers was not only pondering similar ideas, but building actual machines. In a paper titled ‘Zatocoding Applied

---

16 As Segesta and Reid-Green argue, asking ‘if the computer could help’ had become ‘a natural question’ (2002, p. 28) by 1953 already.
17 In an oral history interview, Mooers himself claims primacy for the general idea and argues that Taube’s success came because ‘[h]e was a great salesman and a smooth talker and he charmed the librarians’ (Mooers and Mooers, 1993, p. 13).
to Mechanical Organization of Knowledge' (Mooers, 1951), he described a system using punched cards and the postcoordination logic of independent ‘descriptors’ that could be combined to create complex queries. Interestingly, Mooers’s presentation did not focus on the question of classification but on solving the limitations of punched card systems. Zatocoding was a form of superimposed coding that drew heavily on Shannon's statistical theory of information to encode a much larger number of terms onto the limited number of fields available on a card. This allowed him to build a system using the well-mastered and thus relatively cheap punched card technology, which became ‘the first broadly marketed system that did not use precoordinated (like Library of Congress subject headings) subject terms’ (Williams, 2002, p. 22). While Mooers’s implementation of coordinate indexing worked well with existing equipment, Taube’s combination of reference numbers and lists was almost perfectly suited for the operating principles of digital computers. Experiments with IBM’s first commercial scientific computer, the IBM 701, clearly showed the workability of the method in 1954 (Bracken and Tillitt, 1957; Segesta and Reid-Green, 2002). After all, comparing lists is something that computers do very well indeed.

Overall, coordinate indexing funneled into quite different working systems that could be demonstrated, evaluated, and experimented with. These systems, understood as technical individuals combining coordinate indexing principles with other elements, promised to deliver what Charles Baggage already identified as the machine’s ‘one great advantage’, namely ‘the check which it affords against the inattention, the idleness, or the dishonesty of human agents’ (Babbage, 1832, p. 54). This both resonates with the ‘pessimism about the scope and quality of human reason’ (Heyck, 2012, p. 100) underpinning the overt distrust in librarians and highlights the computer’s capacity to take over simple tasks like finding the same numbers in long lists, which become infeasible for humans beyond a certain point due their repetitive nature and mind-numbing dullness. Coordinate indexing thus reveals how Peters’s (1988) assertion that ‘[t]he computer existed as a practice before it existed as a machine’ (p. 15) is both enlightening and insufficient. The technique shows how older and more elementary principles can find expression in contemporary practices, traveling through different technical trajectories, in this case from notched cards to manual list comparison to digital computers. But coordinate indexing also sits at the precise point where these continuities begin to dissolve into a technical universe where a machine can perform a range of simple operations so quickly and competently that it ‘has outgrown its basic triviality by several orders of magnitude’ (Dijkstra, 1974, p. 608).
A Postcoordinated Universe

The introduction of coordinate indexing is not a story of traditional analog approaches being swept away by the digital computer. The various applications built around it find a certain success, but even in 1975, Van Rijsbergen still emphasizes that ‘[m]any automatic information retrieval systems are experimental’ (Van Rijsbergen, 1975, p. 1). Public libraries have long remained unaffected by information retrieval: when they finally begin computerization in the 1970s, they focus on digitizing existing card catalogs and on making them more easily available as OPACs (Online Public Access Catalogs) (cf. Husain and Ansari, 2006). Boolean operators have certainly found their way into more recent bibliographic systems and simple keyword indexing coexists with more elaborate techniques like subject headings and hierarchical classification. Computerization and digitization can lead to standardization and normalization, but they also provide margins of plasticity, with the effect that actual working systems are often hybrids fusing seemingly contradictory techniques into productive arrangements.

Rather than follow coordinate indexing through historical instances of adoption and adaptation, I want to develop two different lines of interpretation. In Chapters 5 and 6, I will show how the basic ‘term on item’ setup becomes the starting point for a variety of statistical techniques that seek to deepen its empirical dimension, in particular regarding automated indexing and ranking. In the last section of this chapter, however, I want to dive further into the notion of postcoordination. This will lead me beyond coordinate indexing to one of the most influential techniques in all of computing, the relational model for database management.

Postcoordination

The question of how knowledge should be arranged has never been innocent. Darnton (1999, p. 191ff.) explains how Diderot and D’Alembert’s seemingly minor modifications to the tree of knowledge inherited from Francis Bacon was in fact a risky political move that reshuffled the place of secular and religious ideas, and, in doing so, installed the philosophers of the Enlightenment as the supreme guardians of knowledge. Scholars such as Sanford Berman (1980, 1989) have repeatedly argued that traditional but still heavily used classification systems such as the DDC have strong biases, sidelining not only non-Western cultures on the whole but also marginalized groups in Western societies. As Bowker and Star have shown
in their influential *Sorting Things Out* (1999), classification systems have infrastructural qualities that organize and orient important aspects of our lives. The epistemic character of particular knowledge ordering schemes is an imminently political issue, even without taking computers into account.

But the rejection of library principles by early information scientists is not so much a rejection of a particular classification or thesaurus. It is not a battle pitting one taxonomy against another but a fundamental contestation of two thousand years of knowledge organization en bloc. The pioneers mentioned in this chapter never pursued the mechanization of established practices and principles. What they developed and promoted was a new way of conceiving order as ‘ordering’, a dynamic process and practice that was fundamentally different from the creation and application of stable, universal systems. While the information wars have sometimes been framed as pitting ‘people’ against ‘machines’, this narrative obfuscates a more conceptual and indeed more political disagreement about where one should turn when looking for order and organization. What the information scientists propose – and this may sound familiar to readers versed in Silicon Valley disruption talk – is that knowledge ordering no longer needs stewards that guarantee the coherence and adequacy of a tree or ‘world map’. Diderot and D’Alembert’s philosopher, acting as cartographer of knowledge ‘at a vantage point, so to speak, high above this vast labyrinth’ (D’Alembert, 1995, p. 47), is no longer seen as the guarantor of impartiality, but as a problem: overreaching, slow, and subjective. The very ‘image of objectivity’, to speak with Daston and Galison (1992), is under attack.

On a very general level, we find the charge that traditional ordering schemes serve the needs of the library rather than its patrons (Lerner, 2009, p. 189), going back to Thomas Carlyle calling Anthony Panizzi’s catalog ‘a vanity of bibliographical display’ (Svenonius, 2000, p. 10). But the change in perspective is more concretely tied to the real and imagined audiences for catalogs and retrieval systems: the library traditionally adheres to an inclusive mandate emphasizing public education and universal access while the special libraries and documentation centers embrace the specialist looking for actionable answers rather than broad knowledge. If the librarians following the Enlightenment tradition emphasize the creation of a shared, systematic, and stable map of all knowledge that functions as an institutional gateway to a well-tended garden of knowledge, the early information scientists see a growing rift between the ‘finished schemes of the nineteenth century’ (Taube, 1952, p. 167), such as the DDC and LCSH, and a twentieth-century witness to ‘the swirling rush of new literature and new forms of literature’ (Taube, 1952, p. 167). Where the former value
the universality of classification schemes as an asset, the latter see it as a liability that risks cementing outdated worldviews and missing the realities of modern science. Positivist universalists like Otlet and Bliss regarded science and knowledge as singular and unified whereas the actual scientists involved in information retrieval saw a complex, dynamic, and highly specialized research landscape, where established knowledge is constantly questioned and disciplinary boundaries shift. While Otlet (1934) shunned even subject headings as ‘too empirical’ and ‘scattered’ (p. 381) to provide the synthetical and well-ordered perspective the CDU epitomizes, empiricism and fragmentation are precisely what information retrieval pioneers championed, at least to a certain degree. The ‘grand narrative’ of the Enlightenment, emancipation through universal education in a shared space of knowledge, had not so much lost its ‘credibility’ (Lyotard, 1984, p. 37) in this context, but rather its relevance to an application space that defined itself in more directly utilitarian terms. Most of the complicated details of bibliographic control were seen as neither necessary nor helpful to exploit information as a resource for knowledge creation and decision-making. On the contrary, debating whether two editions of a book constitute the same work, to name a common example from library theory, was deemed a frivolous waste of time.

Early information science literature thus brims with doubts about the capacity of librarians (or any other nonscientist) to adequately deal not only with the mass of material but also with the subject matter. The struggle for what Wilson (1968) calls ‘descriptive control’ – the ‘ability to line up a population of writings in any arbitrary order, to make the population march to one’s command’ (p. 25) – was real and (information) scientists sought to wrestle it away from human(ist) knowledge mediators. If we understand the bibliographic chain as composed of source material, ordering scheme (or simply ‘library system’), and searcher, traditional librarianship argued that a lot of expertise was required in the middle to assure the tasks of indexing, organizing, and guiding patrons. The new perspective, however, tried to ‘disintermediate’ the process by buttressing expertise at the two ends of the chain: the specialists writing and self-indexing the material and the decision-makers querying the material were the actual experts and the middle suddenly appeared as a source of delay, incompetence, and subjectivity. Coordinate indexing achieved this goal by two means: on the one end, simple ‘uniterms’ (Taube’s term) or ‘descriptors’ (Mooers’s term) were attributed without order or preconceived relationship (precoordination) and could even be taken directly from the text, eliminating dependence on the expertise of catalogers and indexers (Lerner, 2009, p. 189); and on the other end,
postcoordination allowed searchers to create elaborate queries by combining multiple words in Boolean sequences. The central element, here, is the shift of syntactic and semantic expressivity from the bibliographic language to the query language, meaning that ‘the vocabulary control hitherto deemed essential for indexing purposes could be replaced by additional controls introduced during the search formulation and retrieval processes’ (Salton, 1987, p. 3). The map is replaced by a system laying dynamic pathways into a multidimensional information space. Knowledge organization is reframed as a function of knowledge access. Long before the term ‘disintermediation’ becomes a Silicon Valley favorite and ‘postmodern theorists and artists embrace hyperlinks as a way of freeing us from anonymous specialists organizing our databases and deciding for us what is relevant to what’ (Dreyfus, 2009, p. 12f.), information scientists developed coordinate indexing as a means to orchestrate knowledge organization around an algorithmic technique.

This technique indeed relies on highly queryable information grids that index atomized items with a vocabulary that is itself analytically disassembled into singular terms. The use of Boolean operators already points to the role of calculation as a means for synthetical reassembly. The standard model uses set theory to capture these processes in mathematical terms but coordinate indexing can be conceptualized in more than one way. Although ‘coordination’ refers to the combination of search terms (Varlejs, 2004, p. 96), it evokes a formal or conceptual understanding of a bibliographic system as a ‘coordinate system’, a geometric construct that assigns a place to each item in a multidimensional ‘document space’ lacking a clear navigational form like a list or tree structure (cf. Luhn, 1953). A coordinate index, that is, a space constituted and populated by items and associated terms, represents what I will call throughout the book an ‘intermediate form’, an abstract representation that results from a process of formalization into a normalized structure. Intermediate forms both enable and require calculation to generate specific outputs like result lists or visual diagrams. Creating such outputs represents an act of ordering that is distributed over two components: data adhering to an intermediate form, for example, a coordinate index covering a specific collection of documents, and a technique for algorithmic inference, for example, a system that implements Boolean querying.

---

18 I am generalizing this notion from Tan (1999), who argues that text mining has two broad components. The first, text refinement, basically contains the transformations that yield a ‘purified’ representation of a text, that the second step, knowledge distillation, will then process. This purified representation is referred to as intermediate form.
Whether we think about the search process as intersecting sets of terms or already as a geometric operation is less important, at this point, than the idea that the query indeed defines the category or conceptual ‘location’ of objects to be retrieved. A logical condition – keywords coordinated with the help of Boolean operators – is submitted into a document space defined by an indexing vocabulary and the system returns a specific selection of items that meet the condition as the result. Order, here, is no longer a stable representation, but the momentary outcome of a calculative process that combines indexed items into ad hoc categories. While the card catalogue already introduced a form of mobility that, in practice, ‘rescued’ the stable organization of knowledge by allowing the universalist systems of the nineteenth century to integrate new works into their folds without too much trouble, the mobility of the coordinate index is more fundamental. Unlike the knowledge maps of the DDC or CDU, coordinate indexing is no longer a system of order but a postuniversalist ordering technique that folds knowledge into a specific and temporary arrangement upon request.

This is a first step into a computational universe, even if the technical simplicity and widespread use of what boils down to a combination of tagging and Boolean search makes it difficult to think of coordinate indexing as a particularly powerful algorithmic technique. But Mortimer Taube, although not a mathematician himself, already observed that the mechanical transformations at hand have the potential to provide a genuine epistemic surplus:

[T]he system of bibliographic coordination would, if properly set up, disclose to the searcher more information than had been put into the system. It has become fashionable to state that we can get nothing more out of machines than we put into them. This is like saying that since mathematics is tautological there is never any more in the conclusion than there is in the premises. It is certainly true that when we add a column of figures with an adding machine we get an answer that we did not know even though the answer was implicit in the column of figures. Similarly, in the system of bibliographic coordination, by combining various categories, we may get information which, although implicit in the system, was never explicitly recognized. (Taube, 1951, p. 71)

Although Taube mainly thought about the enhanced potential for information retrieval afforded by Boolean operators, we can easily imagine forms of calculation that describe the document collection on the whole, for
example, in terms of ‘co-tagging’ patterns: if we consider two terms to be ‘related’ if they are applied to the same item, we can generate a network that shows the relationships between all tags used in a collection, providing a conceptual map that could be used as overview or means of navigation. Much of algorithmic information ordering indeed relies on techniques that make ‘implicit’ information explicit in the way Taube describes, often by harvesting expressions of ‘latent human judgment’ (Kleinberg, 1999, p. 606), in this case patterns in term attribution.

The epistemic surplus an algorithmic ordering technique is able to deliver usually springs from the combination of many small instances of counting or calculation, that is, from ‘methods of buying originality with plodding’ that depend on the existence ‘of slaves which are […] persistent plodders’ (Wang, 1960, p. 3). Computerization has made such persistent plodders universally available and datafication has multiplied the ‘items’ available in digital form. Early uses in scientific data processing already indicate that Boolean operators can be applied to any set of entities that can be captured into the intermediate form of a coordinate index. If this requirement is met, the full set of algorithmic manipulations becomes readily available. The considerable possibilities for adaptation and extension become clearer if we consider, for example, the use of the basic ‘term on item’ principle in websites implementing collaborative indexing. While the word ‘folksonomy’, a portmanteau of ‘folk’ and ‘taxonomy’ (Vander Wal, 2007), no longer raises eyebrows, the proponents of free, user-driven keyword attribution, popularized by sites like Flickr or Delicious, not only reiterated many of the arguments19 made in favor of coordinate indexing in the 1950s but showed how multiuser tagging can be processed in multiple ways to produce rankings, recommendations, or semantic enrichment.

With an eye on Simondon’s theoretical apparatus, one could argue that information ordering has developed around trajectories that combine basic principles such as keyword tagging, intermediate forms promoting calculability, and algorithmic ordering techniques into working systems, each element providing ample opportunity for variation. Coordinate indexing constitutes such a sprawling lineage, even if the genealogical connections between contemporary variants and the fundamental propositions by pioneers like Mooers and Taube have largely faded.

But much like the complicated trajectories that constitute what I have called, somewhat callously, ‘traditional’ library techniques, this is not a

19 Mathes (2004), for example, explicitly presents Folksonomies as an alternative to the DCC, LCSH, and other forms of controlled or hierarchical classification system.
linear journey toward some real or imaginary telos. Interestingly, Taube himself tempered expectations concerning the mechanization of coordinate indexing and later became a vocal critic of mathematization, in particular regarding measures of relevance (Taube, 1965). But Taube’s version of coordinate indexing became a (commercial) success and represents an early manifestation of the resolutely perspectivist direction the field had taken.

Shera gives a particularly vivid depiction of that vision:

[T]he classification of knowledge will not be a fixed and unalterable pattern, based on a priori assumptions regarding the most neatly logical arrangement of books on the shelves, but a series of quite widely varying schematisms, each constructed for a specific purpose, or purposes, in accordance with a particular point of view or philosophic orientation. (Shera, 1952, p. 17)

But does information retrieval really subscribe to epistemological pluralism? Is it a harbinger of a postmodern reservation toward ‘grand narratives’ (Lyotard, 1984)? I would suggest a somewhat different interpretation. Taking Hjørland’s (2010) classification of indexing approaches based on ‘epistemological assumptions’ (p. 74) as reference, there is little reason to belief that the perspectivism manifest in coordinate indexing is based on serious hermeneutical or critical doubts about the very possibility of knowledge independent from interpretation. What we find, at this stage, is either a pragmatic empiricism that holds that the world is knowable, even if knowledge necessarily remains partial and guided by interest, or a form of rationalism that suggests ‘that subjects are constructed logically from a fundamental set of categories’ (Hjørland, 2010, p. 74) that can be permuted at will. Even if the practice of keyword attribution relied mostly on empirical principles, controlled vocabularies could tend in the direction of Ranganathan’s analytico-synthetic method which sought to reduce ‘the vast universe of knowledge to a set of atomic concepts and certain basic relations among them’ (Svenonius, 2000, p. 175). While the historical configuration is more complicated, one could argue that coordinate indexing bifurcates into the largely empiricist tradition of information retrieval, which relies predominantly on statistical techniques, and a more rationalistic, logic-based trajectory that is most clearly epitomized by the relational model for database management.

20 Maybe unsurprisingly, Taube’s first customers were the US military and, later, NASA.
The Relational Database Model

Although the relational model for database management, one of the most pervasive technical ideas in all of computing, would easily merit a full chapter or book, even a shorter discussion can help us understand how postcoordination principles have found their way into techniques that define contemporary information ordering more broadly. This is not so much a straightforward case of idea transfer from coordinate indexing to the world of databases but a parallel invention that highlights the contingent character of technical evolution.

The 'library problems' information retrieval set out to tackle were certainly seen as pressing issues, but questions concerning information organization were virulent in other areas of computing, in particular around the different bureaucratic tasks performed in companies and governmental organizations. In areas like accounting, inventory and employee management, payroll handling, or airline reservations, the problem was not to find some document that would respond to a (possibly vague) information need but the precise creation, access, and manipulation of data entries and the generation of reports from the analysis of clearly defined sets of records. Early file management systems placed a first abstraction layer for file manipulation on top of physical storage in the late 1940s and report generators created the kind of aggregate data overviews management decisions could be based on (cf. Haigh, 2006). In the 1960s, we see the emergence of the first Database Management Systems (DBMS) that were not designed for a single task but 'general-purpose' in the sense that users could create their own data models, that is, the semantic structures that would define and organize the files in their databases. These DBMS implemented either a hierarchical (e.g., IBM's Information Management System) or a networked (e.g., GE's Integrated Data Store) database model that defined how users could design their specific data models and interact with the data (cf. Haigh, 2009). Also referred to as 'navigational' database systems, these systems relied on pointers linking one record to the next, making it possible to fast-forward to the correct location on magnetic tape and turning the programmer writing applications on top of a DBMS into a 'mobile navigator who is able to probe and traverse a database at will' (Bachman, 1973, p. 654). Navigational systems were designed to deal with some of the central problems DBMS were facing, namely the speed of data access and the assurance of integrity and coherence when data were modified. But although the pointer system constituted a layer of abstraction on top of the physical storage, the logical representation of data was still largely tied to its physical organization and programmers had
to find pathways through trees or networks of records to get to the desired information. This is precisely the problem that Edgar F. Codd’s famous introduction of the relational model starts off with:

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation).
(Codd, 1970, p. 377)

To achieve a situation where only the ‘natural structure’ (p. 377) of data is visible, that is, the logical or informational structure as specified in the data model, ‘physical data independence’ was a first requirement. But decoupling the logical structure of the data from the idiosyncrasies of the underlying hardware was far from trivial. Coordinate indexing for bibliographic control was adequately served by early computing machinery because it implied little more than intersecting lists of reference numbers and hardly any data modification. Handling large sets of critical company files and frequent, possibly complex updates was an entirely different matter. The first commercially available Relational Database Management System (RDBMS), Oracle, thus only came to market in 1979, nine years after Codd’s initial proposal. Sequential storage such as magnetic tape was not well-suited for the level of detachment the relational model sought to realize and hard disk storage had to become more common to practically succeed in ‘hiding away’ (most) concerns with physical representation from users.

The considerable difficulties in implementing the relational model are in stark contrast to its simplicity on the user side. All data are stored in an intermediate form called ‘relation’ – ‘basically just a mathematical term for table’ (Date, 2004, p. 26) – that has the familiar shape of rows of records sharing the same fields or columns. As Dourish (2017) explains, what is being ‘related’ here are the data items or fields in a row so that ‘one row expresses the relationship between the name “John Smith” and the social security number 123-45-6789’ (p. 117). Unlike coordinate indexing, where all items were covered by the same vocabulary, the relational model allows for the creation of many different tables, each defining its own variables. The requirement that row and column order in a table should be irrelevant (Date, 2004, p. 148) reminds us that relations are again abstract data structures waiting to be queried rather than visual representations already carrying a particular sequence. In coordinate indexing, the terms attributed to an item constitute a set rather than a list, because their sequence carries no importance or meaning. The first indexing term is not more significant
than the last one. Likewise, in the relational model, the grid of rows and columns is a relation rather than a table because the order of rows should not carry any information.

Fields can have a number of predefined types, such as INT (integer), TEXT, BOOLEAN, or DATETIME in the case of the popular MySQL RDBMS. Although the relational model is clearly much richer than a simple term on item system, critics still consider it 'semantically impoverished' (Stonebraker and Hellerstein, 2005, p. 19) because complex data structures, such as classes, are not native to the model and have to be (awkwardly) modeled out of tables. Indeed, to create more elaborate semantic constructs at the data-modeling stage, database designers use primary key/foreign key relationships to connect tables to each other.

Looking at the mock-up data model shown above, one can imagine that cities may appear in all kinds of contexts, for example, as capitals of countries or as locations for monuments. In this case, the table for countries stores two links (foreign keys) to another table, which holds the actual information about cities and attributes a unique identifier (primary key) to each city. The table recording monuments links to both 'countries' and 'cities' in the
same way. If some information about a city changes, it is not necessary to modify each table where city information appears, only the specific row in the ‘cities’ table. This avoids one of the cardinal sins in relational database design, redundancy, and makes data manipulation much less cumbersome. The process of chopping up the world into information atoms without redundancy is called ‘normalization’ and it is part of the data-modeling stage, where the application domain (e.g., cities and their monuments) is captured into a set of tables. Job titles like ‘database architect’ indicate that there is a real body of professional knowledge that informs these practices.

Normalization entails analytical disaggregation to make sure dependencies between attributes are arranged in ways that database integrity and maximum expressivity at query time are guaranteed. Although the relational model provides a richer semantic vocabulary of data types and the possibility to specify precoordinated relationships between tables, modeling practices follow coordinate indexing in refraining, as much as possible, from using composite elements. All information should be distributed over separate and independent variables or ‘coordinates’. The more fine-grained the decomposition, the greater the potential for recomposition at query time. If, for example, an address is properly decomposed into postal code, city, district, street name, and house number, each of these elements can become a means for making distinctions between informational items and, by extension, the things they stand for. One cannot easily list the inhabitants of a city in a database where city names in addresses have not been isolated into a separate field. Normalization is an ‘elimination procedure’ (Codd, 1970, p. 381) that seeks to find and decompose all such instances.

It should thus come as no surprise that the second main ingredient of the relational model, next to the relation as universal intermediate form, is a powerful query language that incorporates further possibilities for postcoordinated selection of items, but also for aggregate description and different forms of ranking or grouping. Here, the main difference with earlier database models is that data is not retrieved by writing a program that forages through a tree or network of records (record-at-a-time principle) but by making a declarative statement (query) that demarcates a group or class of records fitting the statement (set-at-a-time principle), much like in coordinate indexing. Similar to the move from the elaborate structures of the LCSH and DDC to simple keyword combinations, we switch from navigating complex prearranged trees or networks to logical declarations.

21 ‘A relation is in first normal form if ... none of its domains has elements which are themselves sets’ (Codd, cit. in Date, 2006, p. 109).
using a grid of attributes. Based on more advanced predicate logic rather than the basic Boolean logic used in coordinate indexing, query languages such as SQL (Structured Query Language) have considerable ‘linguistic power’ (p. 381) in the sense that they make it possible to submit complex ‘questions’ to a database. Here are some examples in MySQL syntax, the SQL dialect used by the MySQL RDBMS:

```
SELECT * FROM table_name WHERE condition ORDER BY field_name;
```

The basic schema of a MySQL query. In natural language, it would read: select the records, showing all fields ("*"), from the table ‘table_name’ that satisfy a certain condition and order the result table by the field ‘field_name’.

```
SELECT name FROM countries WHERE population > 1000000 ORDER BY population DESC;
```

Here, we are using the basic query schema to select the names of countries with more than one million inhabitants, ordering the results from largest to smallest.

```
SELECT * FROM cities WHERE name LIKE "new%";
```

This query retrieves all cities with names beginning with ‘new’.

These examples show how different field types allow for modes of selection that go substantially beyond the binary presence/absence condition coordinate indexing is built on. Text search, numerical cutoffs, or, in more recent systems, geographical queries are some of the examples for the richer semantics and much more developed calculative expressivity RDBMS allow for. Since the result of a query is also a table, queries can be stacked on top of each other and stabilized into virtual tables, called ‘views’. Queries can also cover several tables at once, using for instance primary/foreign key relationships to draw data together.

There would be much more to say about the relational model, but I hope that my short presentation shows how the principles of atomization (disassembly) and postcoordination (reassembly) have found expression in other areas of computing. While recent years have seen both extensions to the relational model and the proliferation of alternatives for certain task profiles, in particular in areas where retrieval speed primes over query power, it is hard to overstate the importance of RDBMS in the context of almost any bureaucratic process. In line with the postuniversalist perspective outlined
above, the relational model emphasizes a local approach to the modeling of databases and ‘gives no indication about the way in which the world is to be represented by a collection of relations’ (Schmid and Swenson, 1975, p. 212). As a general-purpose technique, it provides no classification systems itself, merely the building blocks for modeling data into atomized units with disassembled properties and the query engine as a means for flexible postcoordination. The goal is ‘application-independent design’ emphasizing decomposition into the smallest logical components in order to remain open for future uses: ‘we are primarily concerned with what the data is, rather than how it will be used’ (Date, 2004, p. 330). Data modeling with an RDBMS has indeed stronger affinities with Ranganathan’s analytico-synthetic and ‘rationalist’ (Hjørland, 2010, p. 74) technique, which seeks to disassemble the world into its smallest logical components, than with the more empirical, pragmatic, and purpose-oriented approach to uniterm selection in coordinate indexing. The difference in broad ‘philosophical’ outlook is one of the ways we can recognize how ‘each kind of information technology embeds its own definition of information’ (Haigh, 2009, p. 20).

From the start, the relational model emphasized combinatorial analysis, that is, the creation of new information through querying, over mere storage and access. Indeed, it explicitly carried the hope that hiding the technical complexities of data storage behind a simple, yet powerful model would turn managers and other nonprogrammers into direct users of database systems, allowing them to model and query the databases that support their daily practices themselves.22 Even if the SQL-fluent business executive probably remains an exception, the relational model has certainly facilitated the integration of databases into organizational processes, supporting informational control not only through much greater ease for storing and accessing information but also through the creation of knowledge layers that establish, examine, and process distinctions in various ways. While terms like ‘knowledge discovery’ or ‘data mining’ point toward processes of examination and ‘exploitation’ that are not necessarily limited to relational data pools, RDBMS have rapidly integrated various types of calculation and analysis, for example, statistical techniques.

SELECT AVG(population) FROM cities WHERE Y(location) < 0;

This example calculates the average population of cities in the southern hemisphere.

22 Driscoll (2012) even argues that RDBMS were central to a wider interest in personal databases and ‘database populism’ in the 1980s.
This MySQL query shows how easily selection and statistical description can be combined, yielding rich and malleable analytical possibilities that allow for complex forms of interrogation and aggregate description that may not have been anticipated when the initial data model was created. These capabilities are particularly interesting in settings where databases store transactional data, for example, sales records or inventory movement. Here, the discovery of patterns or trends can directly inform managerial decisions. This opens onto the vast field of practices and techniques often addressed as 'business intelligence', which rely heavily on the exploitation of atomized data stored in RDBMS. An exploration of the various direct and indirect relationships between the relational model and a ‘new spirit of capitalism’ (Boltanski and Chiapello, 2005) emphasizing flexibility, autonomy, and the capacity to adapt to change is beyond the scope of this book, but it is hardly far-fetched to claim that there has been no technology more central in helping ‘giants learn to dance’ (Kanter, 1989).

This highlights a major difference between coordinate indexing and the relational model: the former was initially applied to ‘a population of writings’ (Wilson, 1968, p. 25), whereas the latter hoped to capture many different kinds of population from the beginning, most significantly populations of human beings. In this context, neither the standardizing and normalizing effects of database practices should be left unmentioned, nor their potential to impose ‘an impoverished, limited language, one that uses the norm to constitute individuals and define deviants’ (Poster, 1990, p. 94). But such forms of bureaucratic and informational control can be exerted with the help of almost any database system. What distinguishes the relational model is its flexibility, which makes it useful for totalitarian modes of power as well as supportive of Deleuzian models of control where real-time modulation primes over static attribution (Deleuze, 1992). It vividly reminds us that customization and standardization are not opposites but form complicated relationships (cf. Star and Ruhleder, 1996). If anything, the relational model makes it particularly easy to move from the individual to the population and back again, allowing for modes of grouping that are only limited by the granularity of the data. If a process of formalization is articulated around atomization, it increases the potential for subtle modes of differentiation and ‘soft’ bio-politics (Cheney-Lippold, 2011, p. 165) where identity and difference are established dynamically and not as hard cuts along preestablished categories.

I will come back to these questions, but to build a broader perspective on the technicity of contemporary information ordering, I want to caution against a perspective that sees databases, and RDBMS in particular, exclusively through a bureaucratic or administrative lens. Their application
space has become much more general. Today, RDBMS form the information storage backends for many different programs and services. Running as server software that application programs can access for their own purposes or distributed in the form of programming modules ready for direct integration, they function as constitutive elements in myriad systems, forming a central core between interfaces for data collection and output facilities. The popular content management system WordPress, for example, has at its center a MySQL database that is fed through an administrative interface. An infinite number of navigable views on the stored contents can be generated with the help of a powerful and flexible template engine that draws directly on the combinatorial possibilities of the underlying database system. Many of the adaptive capacities of web interfaces I have highlighted in the introduction are made possible by the presence of an RDBMS running in the background. Information ordering algorithms, such as ranking or recommendation engines, are performing calculations that decide which story to bring to the top or which video to play next, but this requires that data are ready-to-hand in the form of information atoms that can be accessed and grouped at will. Facebook, to name one of the most impressive examples, uses a complex and heavily optimized data infrastructure to handle the petabytes of information fragments it assembles into familiar interfaces. But this infrastructure still runs MySQL at its core, producing granular and highly queryable repositories of data that ranking and recommendation – including personalized advertisement – can then draw on (Matsunobu, 2016).

Following these operative chains, the next chapter turns to a second tradition building on the coordinate indexing model. Espousing an empiricist rather than rationalist perspective, this tradition makes heavy use of statistical techniques to count a world cut into pieces.

Bibliography


Young, L. C. (2017). List Cultures: Knowledge and Poetics from Mesopotamia to BuzzFeed. Amsterdam: Amsterdam University Press.