The world according to grep: A progress from closed to open?

Paul Duguid[1]

In recent years the Internet has increasingly been defined by search, its resources reached primarily through a search box.[2] While the Internet is new, search of course is not. And though modern search may to appear have shrugged off much of the old apparatus to make information appear increasingly "free" and autonomous, a historical understanding of how that apparatus developed can help clarify what is and is not new and perhaps what is and is not possible for the developing world of digital search. In a highly speculative, partial, and truncated form, this essay attempts to give an idea of what such a history might look like.

The world according to grep

The separation between the short epoch of digital search and its long analogue past can be marked by two "Gs", grep and Google. Grep is the enormously powerful search tool built for Unix software in 1973.[3] It allows searchers to scan digital documents for any "regular expression", any specifiable chunk or "string" (this can include letters, "wild cards", spaces, punctuation marks, or line endings) in the electronic representation of a text. Its closest antecedent was probably the cumbersome concordance, but that was word-based, a limitation left far behind by grep. Google's intervention can be measured by its difference from even its digital predecessors such as Veronica, Archie, Alta Vista, and Yahoo. These relied in varying degrees on hierarchical order for their searches. Google implicitly abandoned that approach, allowing us to hunt down strings across, and

---

1   A version of this paper appeared as "Search before Grep" in Konrad Becker & Felix Stalder, eds., *Deep Search* (Innsbruk: StudienVerlag, 2009).

2   The same, if my own practice is representative, may be true of the personal computer: tools like Apple's "Spotlight" and Google's "Desktop" search tools are making the old, hierarchical "desktop" ordering less important.

3   According the *Netizens* anthology, "grep" stands for "global regular expression print", and was written by Ken Thompson as a search command for Version 4 Unix in November 1973. See Michael Hauben & Ronda Hauben, *Netizens: On the History and Impact of Usenet and the Internet* (1995) chapter 9. Online: Available at http://www.columbia.edu/~hauben/project_book.html. Visited December 22, 2008.

regardless of, the hierarchical boundaries that traditionally organized documents and ranking results instead by the intertexutality of the Internet.[4] Grep and Google, then, brought search into a world of information that was indifferent to semantics, syntax, and hierarchy. Search thus became both a mechanism for and an example of the shift away from ancient constraints, material, conceptual, and institutional, towards the "open" informational environment that is championed by proponents of "open source" and "Web 2.0".[5]

The vision behind open, modular, and hierarchy-free information is not entirely new. Paul Otlet, the grandfather of "information science" long ago suggested that the book was less a resource for information than a constraint upon it, arguing

> The external make-up of a book, its format and the personality of its author are unimportant provided that its substance, its sources of information and its conclusion are preserved and can be made an integral part of the organization of knowledge, an impersonal work, created by the efforts of all ... the ideal ... would be to strip each article or each chapter in a book of whatever is a matter` of fine language or repetition or padding and to collect separately on cards whatever is new and adds to knowledge.

Bill Mitchell, the dean of MIT's Media Lab, expressed a similar view of the book as a constraint on free-flowing ideas when he described it as no more than "tree flakes encased in dead cow"--an outdated old technology ripe for replacement by new. Elsewhere, arguments for "virtual" libraries or "libraries without walls" as well as the simple claim that the Internet is a library, indicated a desire to shake off information-constraining burdens of the past and aggregate knowledge or information, Otlet style. The writer Stewart Brand summed up such ideas with the compelling phrase "information wants to be free".[6]

Whatever the oddities of these claims--Mitchell seems unaware of the incongruity of leather bound books with wood-pulp paper; Brand of whether information is the sort of thing that has wants, let alone is capable of freedom--history, at least the history told by the victors, seems to be on their side. The tools provided by the 2 Gs have

---

4   "Yahoo" stands for "yet another hierarchical officious order", a phrase which sums up the antagonism towards hierarchy and order by even those who felt they depended upon it. See http://docs.yahoo.com/info/misc/history.html. Visited December 22, 2008.

5   For a critique of anti-institutional sentiment about the Internet, see Megan Finn, Daniel Kreiss, and Fred Turner, "The Iron Cage in the Network Society: Some Reminders from Max Weber for Web 2.0," in preparation.

6   Paul Otlet, *International Organization and Dissemination of Knowledge: Selected Essays of Paul Otlet* (New York: Elsevier, 1990), quotation at p. 17; William Mitchell *City of Bits: Space, Place and the Infobahn* (Boston: MIT Press, 1996).

1/16/12

triumphed, allowing us to search in ways never possible before. String search à la grep has allowed us to follow our interests, rather than subordinating them to concordances and hierarchies formed elsewhere. And information-shucking devices such as Google Books have shed material and institutional constraints to provide unprecedented routes to enormous amounts of data, allowing Google, while eschewing hierarchy, to talk nonetheless of "organizing the world's information". Without detracting from the remarkable successes of these technologies, it does seem worthwhile to ask whether, as is often assumed, this progress is part of a continuing movement in the history of search from closed to open, from bounded to free information, from in sum a benighted past to an enlightened future.

A history of the world before grep and Google offers to throw some helpful light on this question. After all, the more we claim that present capabilities are unprecedented, the more we oblige ourselves to study the past, otherwise how do we know what is or is not unprecedented?[7] A glance over the history of search suggests that it takes a fair amount of Whiggish thinking to portray it as the linear emancipation of information wherein the cumbersome shackles of material and institutional constraints were broken by technological innovations and only set back by revanchist assaults on progress.

Such teleological accounts tend to include Darwinian or Spenserian assumptions about human behaviour. Kilgour, a more serious scholar than many who talk teleologically, offers a version of this kind of history in his *Evolution of the Book*. To account for selection and extinction in his evolutionary story, Kilgour's relies on an innate human trait of information foraging. Thus, for example, he argues that "The need to find information more rapidly than is possible in a papyrus-roll-form book initiated the development of the Greco-Roman codex in the second century".[8] I hope to show that such an account of human needs, while beguiling, undersells both the complexity of the past and the challenges of the future. A human need for information is taken in such accounts as an ahistorical, acultural constant, engaged in at all times by all people, who

---

7   In a similar vein, the poet Kipling asked "What should they know of England, who only England know?" See Rudyard Kipling, "The English Flag" in *Writings in Prose and Verse* (New York: Charles Scribner, 1899).
8   Frederick G. Kilgour, *The Evolution of the Book* (New York: OUP, 1998), 5. See also "The need for readily available information, which had been steadily rising, was accelerated by the advent of Christianity", *ibid,* 48.

are also taken to be forever in search of better foraging technologies. While useful in some ways, such an information-driven account can be misleading, describing what we see our distant ancestors as doing, but missing what they saw themselves as doing. Ignoring the gap between these two allows us to enroll the past into an endorsement of present interests, an endorsement which, when actually consulted, the past does not always seem prepared to give. The actual patterns of human behaviour are, I hope to show, more complex than a simple evolutionary or emancipatory account claims in part because the information constraints we so often want to overcome may simultaneously be information resources serving not only to help us find, but also to help us assess, in culturally specific terms, what we find. As assessment is essential if the success of search is not to be random, such constraining-resources often have to be rebuilt in another form that impedes the conventional, progressive story.

Search, storage, organization

To get a grip on grep and Google, we need to look beyond search alone to storage and organization. Google's self-proclaimed mandate, after all, is to "organize the world's information" in order to "make it universally accessible and useful".[9] Moreover, Google's real power probably lies less now in its innovative Page Rank algorithms than in its extraordinary storehouse of open information (refined and tagged by the not-so-open information Google has accrued from searches run over this storehouse); what Battelle has called Google's "database of intentions".[10] The relationship between storage and search is important, because when we look at the distant past, much of what we see is evidence of storage from which we can only infer the historical character of search.[11]

A very early and yet oddly familiar glimpse of storage comes in the opening of the ancient *Epic of Gilgamesh*, which talks of boxes of cedar with clasps of bronze

---

9   See http://www.google.com/intl/en/corporate/, visited December 22, 2008.

10 John Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture* (New York: Portfolio, 2005), chapter 1.

11 Of course, not all storage is done with later search or consultation in mind. The Hebraic *Genizah* manuscripts, for example, were stored simply to prevent the written name of god from being destroyed. Equally, storage is not always done with search primarily in mind. JSTOR, the online database of academic journals, was envisaged to reduce storage costs for libraries. It has become such a powerful resource for searching, however, that its name now seems almost inappropriate. See Roger C Schonfeld, JSTOR*: A History* (Princeton: Princeton University Press, 2003).

holding tablets of stone and lapis lazuli--stone being one of the first means of recording human ideas and different kinds of boxes being, as they still are, useful mechanisms not just for storing, but also for ordering. The relatively valuable materials cedar, bronze, and lapis lazuli then suggest a certain hierarchy of ordering, indicating then as now that the adventures of Gilgamesh were tales worthy of privileged preservation.[12]

In the ancient Middle East, where *Gilgamesh* is set, stone soon gave way to clay. When wet, it is the more malleable material, but when dry almost as durable. Its durability provides our insight into ancient collections of the region such as the "library" of the royal palace at Ebla (c. 2300 BC), many of whose robust clay tablets still survive.[13] These give an idea of what was stored and presumably what was searched. Accounting data, administrative records, and religious incantations dominate this and other early libraries, as well as bilingual word-lists and other forerunners of conventional reference tools. Early collections were relatively small (the room at Ebla was roughly 3.5 x 4 metres). But over this region and the next two millennia collection building became much more ambitious and the content became less pragmatic, shifting from the illiberal to the liberal arts. Great collections, such as the library at Nineveh gathered principally by the erudite Assyrian ruler Ashurbanipal, and the iconic libraries at Alexandria, built under Ptolemy I and II, grew to embrace and give pride of place to philosophy, astronomy, and literary works like *Gilgamesh*. They also grew in absolute terms: the multiple collections at Alexandria held more than 500,000 items.

Ebla's relatively narrow range of documents seem to have been shelved by type and were probably consulted primarily by scribes who created these or similar works. Familiarity would have limited the need for "finding aids". But as collections grew and were consulted by outsiders, the need for more elaborate ordering and finding grew too. The second-millennium collection at Hattusas used a system of colophon marks to identify each record. These suggest some kind of central catalogue for users to find a way to and through particular documents, features echoed on the spines of modern library books. At Alexandria, the first director Zenodotus arranged the collection by type and

---

12 *The Epic of Gilgamesh*, trans Andrew George (London: Penguin, 2003). The paraphernalia were, of course, more indications of hierarchy and canonization--tools of assessment rather than tools for search *per se*.

13 Sites like Ebla would more easily be called archives rather than libraries today, but as Clanchy reminds us, the distinction is a modern one. M.T. Clanchy, From *Memory to Written Record: England 1066-130*7 (Oxford: Blackwell, 1993).

1/16/12

introduced the idea of a catalogue ordered alphabetically. Callimachus, who may also have been a director, later provided Alexandria with a more elaborate catalogue (itself some 120 volumes) that sorted by author and category, limiting the authors to the "eminent" and breaking the categories down into subcategories, and so framing the collection within various hierarchical orderings of the sort still used to manage large, complex libraries.[14]

Immutable mobiles

That such collections could be built from different and dispersed sources (by acquisition or more aggressive appropriation), and assembled and reordered indicates the mobility, adaptability, and to some degree modular self-sufficiency of the works in these collections.[15] In the extreme case, we can contrast this mobility with, for example, the relative immobility of cave paintings, wall carvings, and the like.[16] Such collection building and organizing would have been difficult even with the stones and tablets of *Gilgamesh* and much easier with the papyrus and parchment of Alexandria. Latour has usefully defined documents as "immutable mobiles", and in the transition from stone to the papyrus we see a tension between this pairing emerge.[17] Increasing mobility and pliability, which allowed these great collections to develop and be organized and searched, challenged the immutability that allowed documents to stay constant over time. Stone and clay were resilient in the face of *tempus edax rerum*, but resistant to easy

---

14 For my account of early libraries, I am particularly dependent on Lionel Casson, *Libraries in the Ancient World* (New Haven: Yale University Press, 2001). Even modern hierarchical orders have long roots. The Library of Congress system, widely used in the United States, is based on the system used by Thomas Jefferson, who donated his private library to the country. Jefferson's library was in turn based on a system developed in the seventeenth century by Francis Bacon. See "Francis Miksa, "The Development of Classification at the Library of Congress" Occasional Paper 164, University of Illinois, Graduate School of Library and Information Science, Champagne-Urbana, 1984.

15 The great early collections were built not only by lavish patronage of arts and science, but also by an enduring library tradition of conquest, pillage, and other kinds of forced appropriation. The emissaries sent out to track down acquisitions by fair means or foul are perhaps early antecedents of Google's crawlers, and the libraries, such as those of Babylon and Pergamum, dissolved under this kind of appropriation, the forerunners of dead links.

16 Armando Petrucci, *Public Lettering: Script, Power, and Culture* (Chicago: University of Chicago Press, 1993)

17 Bruno Latour, "Visualization and Cognition: Thinking with Eyes and Hands," *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6 (1986): 1-40.

organization. Papyrus and parchment, by contrast, could be organized, rearranged, and stitched together with relative ease, but they were more easily damaged, intentionally or unintentionally. The contents of Sumerian "tablet houses", much of it the equivalent of modern ephemera, have survived remarkably intact for 5,000 years; the contents of Alexandria have all but disappeared.[18]

Despite increasing frailty, the emphasis on mobility seems to have predominated for all but monumental inscriptions. Texts moved from stone and clay to more amenable material. What material was actually used was in part a function of place. Alexandria took advantage of the plants that grew in the Nile valley to make papyrus. In Greece and Rome, where papyrus was not available, parchment (which takes its name from the great library at Pergamum) or vellum was the main support for library documents, though more ephemeral writings used wood and wax. In India, birch bark was used in the north and palm leaves in the south; in China, boards, bamboo, and silk. Physical geography and plant life were not the sole determinant. In India, for example, the status of cattle ruled out parchment.

China's invention of paper, which can be made more or less anywhere and violates no widespread taboos, ultimately overcame all the competition as material support for all but a few documents. It endures as a type of support to this day (sometimes even used as insurance against the frailty of digital storage), though in terms of its material endurance, it is perhaps the most mutable of all the materials discussed so far except for wax, suggesting that mobility may trump immutability in Latour's dyad.[19] If printing began, as some suggest, with the Chinese tradition of copying Confucian classics by making paper rubbings from their stone engravings, that process also captures the symbolic transition of communication from the primarily immutable to the primarily mobile. Paper, which could be marked and amended, glued, stitched, and appended more easily than most of its rivals advanced possibilities for storing, ordering, and indexing. Its introduction should not be read as a simple advance on information's linear progress

18 Though probably not, as is usually assumed, in the famous fire. See James Raven, ed. *Lost Libraries: The Destruction of Great Book Collections Since Antiquity* (London: Palgrave, 2004).

19 It is always tempting to argue that the frailty of new kinds of support, such as for example digital documents, will prevent them supplanting the old. Trithemius makes such an argument, using the frailty of paper to suggest that print will not be able to compete manuscript. In fact, despite the need for an immutable text, lability often trumps rigidity in such confrontations. See, Johannes Trithemius, *In Praise of Scribes*, trans R. Behrendt (Lawrence, KA: Coronado Press 1974), first published 1492.

towards Brand's freedom and autonomy. The salient features of paper suggest that much of its attraction came because it was particularly well behaved within institutional collections, which in turn could protect its fragility.[20] Paper undoubtedly helped underwrite more powerful search, but within, rather than in isolation, from institutionally based organizational hierarchy.

Bad punctuation

The significance of this new material base for search may in part reflect the way paper helped underpin a new form for documents, the codex, the object of Mitchell's scorn. What we think of today as the modern book, with its stiff cover and sequential, individuated pages folded from larger sheets into signatures, gradually replaced the scroll more or less as paper replaced parchment. As these two seem to belong together (for paper folds more easily than papyrus or parchment, which are better adapted to the roll), it is easily to tell stories about a new technology replacing the old because of its inherent suitability for the underlying human trait of information foraging. Kilgour, who makes such an argument, explains away the long gaps in this process of replacement in which nothing of technological interest happens with the help of Gould's notion of punctuated equilibrium.[21]

Such stories of supersession, extinction, and equilibrium need treating with caution. Too often to make their case for the new and adaptable, they belittle the older technologies as primitive and static.[22] The scroll was in fact a highly adaptable form within which some of the most enduring features of the apparatus of search--apparatus

---

20 Different traditions made different assessments of paper's frailties. Islamic bureaucrats adopted paper quite early, whereas European chanceries were more suspicious and slow to adapt. See Pierre-Marc de Biasi & Karine Douplitzky, *Le Saga du Papier* (Paris: Adam Biro, 2002) for Islamic enthusiasm, Lucien Febvre & Henri-Jean Martin, *The Coming of the Book: The Impact of Printing, 1450-1800* (London: Verso, 1984) for Western hesitation.

21   These gaps usually signal that no major new technology was introduced, which, in technologically determined stories, means that nothing of interest occurred. Such accounts have to leap from paper to print and from print to the steam press, the camera, or the telegraph. For the technologically driven, the twelfth or the eighteenth century, for example, seem to be periods in which nothing happened.

22 This is a regular feature of accounts that stress the advances of the new. Hence Grafton complains of Elizabeth Eisenstein belittling scribal culture in order to enlarge the effects of print. Anthony Grafton, "The Importance of Being Printed," *Journal of Interdisciplinary History* 11 (2) (1980): 265-286.

we tend to associate with the codex and print--developed.[23] In terms of adaptability to human needs it was, moreover, particularly handy, more suited when closed to the human grip than the codex. It was thus easily carried--and concealed, leading Socrates to tease Phaedrus about what he had under his cloak (a question that also teased Derrida). For sure, the codex unquestionably had advantages. It could import many of the features of the scroll, while adding features previously unavailable. It allowed, for example, writing on two sides of a document (as, of course, had clay tablets). It also gave the page (and the two-page spread) a more robust semantic role, and it provided an unambiguous margin, essential for concise annotations and reference markings.[24] While it was perhaps not so easy to carry, it stored and stacked more easily, suggesting again how the needs of storage and organization might take precedence over individual access. Though, unlike the scroll, the small codex or *membrana* could, as Martial argued, be held and read with just one hand (a possibility which teased Rousseau).

Even as we weigh advantages and disadvantages, it is important not to jump to simple evolutionary conclusions about survival and extinction. Given their slightly different properties and potential, it is not surprising that these two forms, the scroll and the codex, existed side by side, as illustrated by a famous painting from Pompeii in which one of the two figures holds a codex and the other a scroll. Nor was this overlap particularly brief. The original codex of wax and wood (from whence its name) was around as a convenient notebook long before it rose to prominence as a high cultural object, threatening the scroll. Moreover, as Clanchy reminds us, long after the codex was widespread, the English adopted the scroll for legal and government documents and it survived as the principal means of recording, storing, and organizing chancery court proceedings until at least the late 19th century. Nor, as screen documents remind us, is it quite dead yet. Indeed, in Google's book project, the pages of the scanned books helpfully scroll, whereas in other more literal translations, such as Early English Books Online, they jump from page to page.

---

23 Apart from colophons and *sillyboi* and the developments of Zenodotus and Callimachus, the scroll introduced markers for the beginning and end of significant textual chunks (*incipits* and *explicits*) and other important internal features.

24 For the importance of marginalia, see William H Sherman, *Used Books: Marking Readers in Renaissance England* (Philadelphia: University of Pennsylvania, 2008).

Casting far and wide

Even if we can see the print book emerge triumphantly to supersede the manuscript scroll, it hard it is to make easy evolutionary sense of the development of the book in terms of information foraging, as even the following survey, geographically as broad as it is shallow and historically as deep as it is lacking in profundity, can show.[25] And once we look beyond Europe, the often-told story of Western triumph becomes increasingly confusing. If we take paper alone and ignore hints that it may have been around for up to 300 years before its official Chinese birth in 105 AD, what seems to some a principal information-foraging resource takes a surprisingly long time to spread to societies that evolutionary models assume were no less-information obsessed than China. It took 5-600 years to reach India and the Middle East and a further 500 to cross the short distance from there to Western Europe. China had also developed xylography by the eighth century, and moveable type by the eleventh, yet, despite the codex and paper together forming a robust manuscript culture in Europe and Byzantium, it still took until the fourteenth century for xylography to develop and the fifteenth for the apparently transformational appearance of moveable type. Moreover for all its info-precociousness, China for its part proved resistant to the pure, search-friendly codex, relying on the "sutra fold" until well into the seventeenth century.[26]

Korea and Japan, though both heavily influenced by and influencing China,

---

25 Along with works specifically cited, the following sections borrow extensively from Thomas Francis Carter, *The Invention of Printing in China and its Spread Westward*, 2d edition, ed. L. Carrington Goodrich, (New York: Roland Press, 1955); Guiglielmo Cavallo, "Du Volumen au Codex: La Lecture dans le Monde Romain," in Guiglielmo Cavallo & Roger Chartier, eds., *Histoire de la Lecture dans le Monde Occidental* (Paris: Éditions du Seuil, 1997), pp: 47-78; Roger Chartier, *The Order of Books: Readers, Authors, and Libraries in Europe between the Fourteenth and Eighteenth Centuries* (Stanford: Stanford University Press, 1994); Simon Eliot & Jonathan Rose, eds., *A Companion to the History of the Book* (Oxford: Blackwell, 2008); Alexandra Gillespie, *Print Culture and the Medieval Author: Chaucer, Lydgate, and Their Books, 1473-1557* (Oxford: Oxford University Press, 2006); William A. Graham, "Traditionalism in Islam: An Essay in Interpretation," *Journal of Interdisciplinary History* 23(3)(1993): 495-522; Paul Lemerle, *Le Premier Humanisme Byzantin: Notes et Remarques sur Enseignement et Culture à Byzance des Origines au Xe Siècle* (Paris: Presse Universitaires de France, 1971); Library of Congress, *Papermaking: Art and Craft* (Washington, DC: Library of Congress, 1968); Malcolm Parkes, "The Influence of the Concepts of *Ordinatio* and *Compilatio* on the Development of the Book," in J.J.G. Alexander & M.T. Gibson, eds., *Medieval Learning and Literature: Essays Presented to R.W. Hunt* (Oxford: Oxford University Press, 1976), pp: 115-141; Francis Robinson, "Technology and Religious Change: Islam and the Impact of Print," *Modern Asian Studies* 27(1)(1993): 229-51.

26 This is a hybrid between the scroll and the codex, whereby pages fold on both their front edge and back edge, rather like a paper blind, so that, though such books fold flat like the codex, when open they are continuous like a scroll.

also reveal distinct chronologies. Korea had paper by the third century, print by the eighth, and moveable, alphabetic type 50 years before Gutenberg, yet the full panoply of the Western print codex and what Febvre and Martin see as a key information device, the newspaper, waited to be "introduced" by the Japanese towards the end of the nineteenth century. Meanwhile, Japan itself might almost appear to have put evolution into reverse. It got paper from Korea early in the seventh century and was capable of printing the celebrated charms of the Empress Shōtoku in an edition of perhaps one million copies in the eighth century, yet printing for reading (the charms were a ritual rather than a communicative act) failed to develop until the eleventh century (well before Europe). Even then was generally ignored until typography reappeared with the Jesuits in the seventeenth century.

Indian palm-leaf books or *pothi* may have inspired the sutra fold used in China. India also had paper from the sixth century. Yet paper did not become widespread for another 700 years, while, despite the use of print by British in India, print was barely used by Indians for themselves until the late nineteenth century. The Indian nations had thoroughly sophisticated mechanisms of communication, as Bailey points out in *Empire and Information*, consequently explanation must lie "in the information order as a whole, rather than in one particular dimension of it".[27] Again, Islamic societies had the codex almost from their beginning, and paper by the ninth century. It was from the Islamic Middle East that paper spread slowly to Europe and Byzantium (in the latter, paper was named after its source, "Baghdad"). Moreover, though Islamic communities often had in their midst both Jewish and Christian printing centres, print did not spread through Islamic nations until the late nineteenth century. For their part, Jewish communities do not seem to have adopted the codex until the ninth century being until then identified, at least in the eyes of Christians, by their affinity for the scroll.[28]

This historical accounting, however inadequate, makes it hard to argue for some kind of simple technological determinism or a fundamental information imperative in human nature. Rather, it suggests that the spread of communication technology is subject to many forces. It is, after all, widely acknowledged that the codex spread in the

---

27 C.A Bayly, *Empire and Information: Intelligence Gathering and Social Communication in India, 1780-1870* (Cambridge: Cambridge University Press, 1998), 239.

28 Emile G.L. Schrijver, "The Hebraic Book" in Eliot & Rose, *A Companion*, pp: 153-164; Lemerle, *Premiere Humanisme*, Clanchy, *Memory*.

West less on the back of its communicative or search capabilities, than as a mark of religious affiliation: It spread with Christianity because Christians used it to distinguish themselves from the scroll-using religions that came before it. Much like the iphone and the ipod, the codex was a cultural marker as much as an efficient device to provide access to information.

It would be wrong, of course, to deny that the codex became enmeshed in sets of practices that look increasingly information-centered and technology driven to us. But it is important not to separate those practices from a vast array of others that do not reduce so readily. In the West the codex was a religious instrument for a long time and its development as a technology for search, organization, and storage reflected changes in the way in which texts were used in that particular confession. Developments in Christian scholarship changed the appearance and use of the codex in the long and often underestimated period between its initial spread and development as the primary medium of print in the West. In particular, in a remarkable period from the eleventh to the fourteenth century, Christian scholars devised new textual and intertextual resources to further access, search, and reference. These advances might almost be thought of as paving the way for print. Driven in part by the spread of minuscule writing in both Western Europe and Byzantium, word divisions, paragraph markers, and punctuation began to break up the text itself into more accessible chunks. Paratextual apparatus also developed, including the increasingly sophisticated gloss, running heads, shoulder notes, the table of contents and the alphabetical index, and the page number. The printed book inherited all these, though some, like the page number and the shoulder note, only with difficulty.[29] Indeed, in the early period of print, search capability does seem to go backwards, suggesting that it was less of an imperative than scholars like Kilgour or Eisenstein imagine.

---

29 Margaret M. Smith, "Printed Foliation: Forerunner to Printed Page-Numbers," *Gutenberg-Jahrbuch* 63(1988): 54-70.

1/16/12

Mutability, reliability, and verification

In trying to understand changing search, we need to note that in the last centuries of the manuscript era in Western Europe, the social context of the book was changing too. Both production and consumption were leaving the controlled confines of the monastery for the wilder terrain of cities and universities. In the growing professions of the book trade new centres of production arose outside the enclosed scriptoria and new readers arose outside the cloistered library. It was more in these new hands than in the old monastic ones that the new tools for search described above developed. But to go beyond simple finding tools and to help searchers assess the worth of what was found, these new sites had to address again the tension--this time of their own making--that emerges when increasing mobility of texts and textual production threatens the stability or, in Latourian terms, immutability of the text. As book production and consumption spread, the threat came not so much from individual documents decaying, but in the variation introduced between versions of the "same" text by proliferating copies.

Openings for this kind of change and corruption were numerous. On the one hand, there was legitimate change. St. Bonaventure famously distinguished four kinds of copying. The lowest involved no more than verbatim repetition of an original, but higher levels of scribal practice permitted the addition of comments from other writers, and from the scribe himself, who ultimately is acknowledged as a new author.[30] On the other hand, there were illegitimate changes, some through incompetence, and some through different kinds of forgery and falsification. Even monasteries, as Clanchy shows, were forced to forge, sometimes producing false forgeries and sometimes true ones, a distinction that complicates the mix even further.[31]

In Western Christian culture in the centuries before print, Cavallo and Stock argue, the codex developed into an authoritative form at the same time as its centres of production were spreading and its audience was growing. The mutability that came with new places and processes of production challenged the book's potential authority.

---

30 Gerald Bruns, "The Originality of Texts in a Manuscript Culture," *Comparative Literature* 32(2) (1980): 113-129.

31 Guglielmo Cavallo, "Du Volumen au Codex"; Brian Stock, *The Implications of Literacy: Written Language and Models of Interpretation in the Eleventh and Twelfth Centuries* (Princeton: Princeton University Press, 1983). The varied practices of this period led in good part to the development of the skills of document-fraud detection known as diplomatics and the remarkable work of Mabillon.

Mutability--in this case less within copies than among them--presents readers, and especially new readers, with a challenge. For if a reader goes to a book searching for new ideas but without a strong background in the domain, he or she is unlikely to be able to judge reliability.[32] A kind of market for intellectual lemons may well develop.[33] A retrospective view suggests that the conflicting challenges of making works findable and making what was found reliable resulted in moves, as they might be described today, to "free" information being countered by moves to restrain it. The world of books found various ways to do this.

Islamic tradition found one solution. It did not take books, in particular religious books, as autonomous. Rather books drew their authority from particular teachers, who in turn were warranted by a "golden chain" that connected them to Muhammad.[34] Elsewhere, books developed more inherent authority, provided not by the text on its own, but by supervening institutions. Again, Alexandria provides an early precedent. Zenodotus took it as part of his role as librarian to produce exemplary texts.[35] Later, the Christian church worked to standardize its texts and remove apocrypha of one sort or another. So doing, the church also spread the carolignian minuscule as a standard script.[36] And universities, first in Byzantium and later in Western Europe also took responsibility for textual integrity of celebrated works. In tenth-century China, where corruption resulting from ease of print and paper was directly contrasted to the reliability imputed to the immutability of stone, the National Academy took up the task of quality control.[37] A related two-step can be perceived with the advent of print in the West. As

---

32 This is a central challenge presented in Plato's *Meno* and *Charmides*.

33 George A. Akerlof, "The Market for Lemons: Quality, Uncertainty, and the Market Mechanism," *Quarterly Journal of Economics* 84(1970): 488-500. The book world Johns describes would seem to embrace several such markets. See Adrian Johns, *The Nature of the Book: Print and Knowledge in the Making* (Chicago: University of Chicago Press, 1998).

34 Consequently, in Islamic cultures, biographies establishing personal connection were a particularly important form, whereas in the West the provenance-establishing bibliography was perhaps more significant. Michael Albin, "The Islamic Book" in Eliot & Rose, *A Companion*, pp: 165-176.

35 I have argued elsewhere that libraries have more to do with quality than they usually own to. See Paul Duguid, "Inheritance or Loss: A Brief Survey of Google Books'," *First Monday* 12(8) 2007.

36 Michelle P. Brown, "The Triumph of the Codex: The Manuscript Book before 1100" in Eliot & Rose, *A Companion*, pp: 179-193

37 Carter quotes one such argument:

    During the Han dynasty, Confucian scholars were honored and the Classics were cut in stone. ... In T'ang times also stone inscriptions containing the text of the Classics were made in the Imperial School. .... We have seen, however, men from Wu and Shu who sold books that were printed from blocks of wood. There were many different texts,

Johns has argued, the reliability of the text, even within editions, was not a function of print alone. Here as elsewhere the institutionalization of publishing acted as a centripetal force providing reliability to counteract the centrifugal tendencies that came with the increasing mobility introduced by the new technologies.[38] Censorship and later copyright were heavy constraints, but they were excused at the time, sometimes with justification, as ways to secure reliable copy.[39]

Encircling and uncircling

Such institutionalization helped conceptually to circumscribe the text within the book, which was in turn within an institution such as the library or similar system of authorization. As we have seen with *Gilgamesh*, combinations of material and institutional constraints, from cedar boxes to the library at Nineveh, are bought to bear on the central texts of a society. This concept of circumscription is implicit in two important scholarly words. One is *encyclopedia*, which invokes an encircled body of knowledge essential to education. And the other is *search* itself, which like *encyclopedia*, is etymologically descended from classical words for a circle and whose roots suggest an encircled body of knowledge to be examined. Search, that is, demarcates not just the needle, but also the haystack. The idea of a corpus that can be encircled and searched implies (and simultaneously deprecates), of course, a secondary body that is excluded as false, forged, ephemera, and the like.[40] Such an idea is physically inscribed in some of

        but there were among them no orthodox Classics. If the Classics could be revised and thus cut in wood and published, it would be a very great boon to the study of literature. (Carter, *Invention*, p. 70)

38 Some early printers signed their output as an attempt to indicate authenticity. Some print shops allied themselves with scholars. See Johns, *Nature of the Book*; Elizabeth Eisenstein, *The Printing Revolution in Early Modern Europe* (Cambridge, UK: Cambridge University Press, 1983). For the complex task of quality assurance in early publishing, see Paul Duguid, "Brands in Chains," in Paul Duguid & Teresa da Silva Lopes, *Trademarks, Brands, and Competitiveness* (London: Routledge, forthcoming 2009).

39 Loewenstein sees this as a "founding myth" of intellectual property systems, but like most myths it contained a grain of truth. Joseph Loewenstein, *The Author's Due: Printing and the Prehistory of Copyright* (Chicago: University of Chicago Press, 2002), 252. For the connection between early forms of copyright and control in France and their relation to the reliability of the text see Mark Rose, *Authors and Owners: The Invention of Copyright (*Cambridge, MA: Harvard University Press, 1993) and Elizabeth Armstrong, *Before Copyright: The French Book-Privilege System 1498-1526* (Cambridge: Cambridge University Press, 1999).

40   Such an idea lies behind Caliph Omar's famous edict that if what a book says is not already in the

our main institutions of learning, in particular, as Chartier points out, in the round reading rooms of major national libraries (from the old Bibliotheque National in Paris and the old British Museum reading room in London, to the Library of Congress in Washington, and Asplund's great national library in Stockholm). These evoke the circles of established knowledge in which the learner can safely and reliably search.[41]

But, like the *salons des refusées* formed by artists that the official art exhibitions of Paris had rejected, such attempts to bound information inevitably engender attempts to break the bounds. It may be chance, but it is indicative that Wikipedia kept the *pedia* aspect of its source, but rejected the encircling or enclosing part. Wikipedia's vision is of open and unrestricted contributions. Symbolically, at least, this speaks to a desire for search not to be circumscribed, as it had been in the past, but to be ever more open.[42] As I have tried to suggest, this impetus is far from new. The path to openness may in fact be more cyclical, than linear. Attempts to break out in the name of freedom lead in turn to attempts to constrain in the name of quality which in turn lead to new breakouts. Another imprudent canter across changes in England towards the end of the seventeenth century allows me sketch a set of prior profoundly important attempts in politics, in business, and in science to break established boundaries of search.

Habermas indicates the change in politics in his account of the development of a "public sphere".[43] The growing bourgeoisie came to believe that citizens, by shedding their particular, personal interests and entering a sphere of open debate, could make the best political decisions. The search for political solutions now turned not to the monarch for answers but instead to free enquiry, rational debate, and the open exchange of

---

Qu'ran, then it is heretical and the book must be rejected; and if what it says is already in the Qu'ran, then the book is redundant and can be rejected. Phrases like "useful knowledge" or "useful information", which are endemic in discussions of collections of one sort or another, similarly acknowledge the idea of boundedness. What is not useful is left beyond the pale of institutions and collections.

41 Chartier, *Order*. This argument owes much to my colleague Geoffrey Nunberg's insights into the place of information.

42 Elsewhere, I have argued that this openness loses for Wikipedia some of the useful resources that come from boundedness. See Paul Duguid, "Limits of Self-Organization: Peer Production and the 'Laws of Quality'," *First Monday* 11(10) 2006.

43 Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (Cambridge, MA: MIT Press, 1989). Habermas's account is undoubtedly idealised and his periodisation problematic (see, for example, Craig Calhoun, ed. *Habermas and the Public Sphere* (Cambridge, MA: MIT Press, 1996)), but his claim that a significant change in politics and political debate took place around this time remains reasonable.

information. This type of search could lead not to some preordained answer, but rather, as it did with the American revolution, to a conclusion that was previously unimagined and perhaps unimaginable--an answer that was in some ways an emergent property of the process of search itself. As such, search did not involve, as it does in Socratic dialogue, uncovering something the interlocutor knows already but is not aware of; and it did not involve, as it did in most scholastic or religious enquiry, returning to what had already been written and revealed. It involved finding what had hitherto been unknown and perhaps even unknowable.[44]

In the same period, the development of stock markets produced a comparable process in the realm of commerce. The search for the price of shares and by extension the value of companies created that price in the very practices of the exchange. Price in such conditions was hitherto unknowable; only within the economic activity of the market did it emerge. By extension, and in opposition to the mercantilists of the period, free-traders suggested that the worth of the nation as a whole could only be determined by openly trading rather than by hoarding (and counting) its gold reserves. Finally, seventeenth century science can be seen as rejecting deference to prior authority and developing a similar, open-ended kind of search. Galileo, Descartes, Boyle, Hooke, Huygens, and the other early modern empirical scientists, to the dismay of figures from the pope to Hobbes, interrogated nature with odd devices and open-ended searches. Science was no longer the province of scholastic figures with ancient books, but open enquiry unfettered by institutions and claims of expertise. Experience, as Chaucer's Wife of Bath had memorably claimed, and not authority was the currency of scientific endeavour.

These extraordinary revolutions broke down many of the prior bounds of search and pointed to an open landscape, addressing desires for individual freedom and distaste for existing institutions and their hierarchies of knowledge. But the story did not end there. The public sphere, as Habermas noted, was transformed and subordinated to bourgeois interests. It took institutional innovation and experimentation in different kinds of democracy to maintain the very notion of openness to which institutions had previously been thought to be anathema. Similarly, the developing markets and their

---

44 In this regard, *search* and *learning* can be ambiguous in similar ways. Both can point to what the searcher/learner did not know before but others did, or to what no one knew before.

open trade met their first global shocks with the South Sea Bubble and tulip mania. Like the public sphere, they came to rely on institutional boundaries that the idea of a free market and an open search for value still holds in contempt. Most markets continue to require institutional intervention and adjustment to ensure that prices are more open than fixed. And science, as Shapin and others have shown, slowly institutionalised as well.[45] Over time, the Royal Society and its remarkably open publication the *Philosophical Transactions* became increasingly closed.

While holding admirable aspirations to freedom and openness, politics, business, and science circled the wagons, limiting legitimate search to certain endeavours, certain methods, and certain types of question, and ruling others out as unacceptable. Dissenting voices were spun off into the realm of alchemy, antiquarianism, cabalism, necromancy, and the like. Indeed it is noticeable that many of the major encircling institutions--from the modern newspaper, the Bank of England, and the Royal Society, to modern encyclopedias were in fact constructed around the time of--and it seems plausible to suggest in response to--these three revolutions in search. And much of the success of their searches was due to the constraints that bound them. It is unsurprising, then, that succeeding centuries mark new attempts to break restrictive bounds and new countermoves to reign in unstructured behaviour. Inevitably, some of the new constraints are imposed by forces of reaction, like Catholicism's response to Galileo. But not all. Some involved the building of Weberian institutions to provide the very conditions for idealized Habermassian discourse, Hayekian markets, or open science in their search for answers.

Let me take one final example from a domain that developed many powerful search tools, analogue and digital, the law.[46] The common law is in many ways an archetypal open system that over time went closed with the development of statutory law. Yet even this process is not linear. Cycles of openness and closedness are evident, for example, in the development of law reports. These published the significant case law on which courts and litigants relied for precedent. At the beginning of the nineteenth century, new printing technologies and the resulting drop in the cost of publishing led to a

---

45 Steven Shapin, *A Social History of Truth: Civility and Science in Seventeenth-Century England* (Chicago: University of Chicago Press, 1995).

46 Mead Data, for example, forerunner of LexisNexis, played an important role in the history of modern search.

bustling market for law reports. The Law Amendment Society noted at the time "it has long been considered a practicable scheme for any barrister and bookseller who united together with a view to notoriety or profit, to add to the existing list of law report". This expansion, the *History and Origin of the Law Reports* noted in the mid-century, resulted from "applying the principle of competition to correct the evils of prolixity, delay, and expense incident to the [old] system of authorized reporting". The process might be described today as having "opened" law reporting, allowing the market to work and superior contributions rising to the top. But such markets for knowledge often work the other way. With proliferating reports and no institutional standards for reporting, lawyers and judges became increasingly worried about the reliability of the reports. While old constraints were broken, the change led to "new evils [that] created confusion and uncertainty in the law". Open competition, championed in the nineteenth century much as it is today, served the interests of individual lawyers and publishers in expanding the production of legal information, but made search and the assessment of what was found more and more problematic as it was "carried on without regard to the interests of the profession or public" and produced "perplexity in the administration of justice". [47] Trying to balance conflicting imperatives, the Law Amendment Society, a reforming organization, managed to institutionalize the system and rein in the proliferating reports without entirely stifling innovation in reporting. Such processes continue to be cyclical, however. New technology has opened the system again, and consequently forward-looking lawyers are beginning to sound much like their nineteenth century counterparts: "The almost universal view among judges in England is that too much, rather than too little, is reported".[48]

---

47 W.T.S. Daniel, *The History and Origin of the Law Reports: Together with a Compilation of Various Documents Shewing the Progress and Result of Proceedings ...* (London: W. Clowes & Sons, 1884), quotations pp 23-4.

48 H.W. Arthurs, "A Lot of Knowledge is a Dangerous Thing: Will the Legal Professions Survive the Knowledge Explosion?" *Dalhousie Law Journal* 18(2) (1995): 295.

Beyond the 2 Gs

As we look at modern search tools, it is easy to believe that they are contributing to a historical march away from closed and restrictive institutions to democratic openness and that amassing more information, regardless of source and running grep like searches across it is inherently a good thing. From beginnings in open source software, this new idea of openness has spread to contemporary aspects of politics, of markets, and of science, as well of cultural endeavour. To speak against it seems hopelessly reactionary. As I have tried to argue, attempts to impose structure are not always attempts to return us to the past. They are often (though again not always) attempts to control what has become, like the nineteenth-century law reports, unmanageable information, information that baffles search not because it is resistant to search algorithms, but because it is not structured in any openly accessible way and hence its findings are inscrutable or unreliable for those who want to use them. It is easy to believe that new technology replaces old institutions. Google, it is often thought, will replace the library. But technologies and institutions are not the same thing. Technology alone is often incapable of imposing useful structure and to increase reliability, but requires complementary institutions that are open and available for public scrutiny. Thus a history of search, as I have tried to illustrate, looks less like a linear progression driven by an innate appetite for information foraging than a set of almost unfathomable cycles around closed and open structures.

Indeed, and I apologize to anyone who has come this far for saying this so late, we can see these cyclical movements without undertaking such a tortuous historical journey. Wikipedia, for example, which has exemplified for many the accrual of open, searchable, and reliable information about the world, has for some time been engaged in building structures. Its early contempt for credentials and expertise and its resistance to hierarchical exclusion has given way to an understanding that claimed experience is not necessarily better than earned authority. Indeed, Wikipedia itself has started to build an institutional structure through its foundation, which is inevitably imposing constraints on the project's openness.[49] Implicitly, the *encyclo* is returning to fence around the *pedia*. More intriguingly, Google has tried to challenge Wikipedia with its Knol project, which

---

49 I'm grateful to Mayo Fuster Morell for conversations that have helped me understand tensions around the Wikimedia Foundation.

pays far more deference to institutions and their role in the creation knowledge, aggregation of information, and the hierarchical ordering of ideas. Meanwhile, in the face of uncertainty as to its quality, Google's own technology-driven library project is being corralled by the Hathi Trust. In this case, attempts to make search more reliable, are being tempered by one of the oldest of contemporary institutions, the university.[50]

---

50 Even open source software, the exemplar of the new spirit of openness, turns out not to be the linear example of the closed going open after all. A good deal of software was quite open by contemporary standards in the 1950s before corporations enclosed it. And much of what is open now is held open by the institutions of the law (which protects "copyleft"), by the self-interest of quite hierarchical organizations from Red Hat to Sun and IBM, and by the university. For software that was open before it was closed, see Michael Schwarz & Yuri Takhteyev, "Half a Century of Public Software Institutions," in preparation.