Chapter 5: Salient Features of Digital Textuality

With the Internet, but more particularly with the World Wide Web in the 1990s, a wholly new textual medium sprang into being. In the previous chapter I approached the birth of this new, digital, medium from a longer historical perspective. Some landmarks in its development truly stood out. After the computer became a Universal Machine it could run every conceivable type of application—of which word processing and networking were indispensable for the function it has now as a textual medium. Word processing was the final outcome of a longer process of harnessing the computer to deal with text at all. As a corrollary of that process, the computer's interface also became primarily textual. That computers next became linked in a global digital network changed their medial role from that of being mere aids in analogue text production and printing to that of a wholly new digital medium in its own right. This new medium had properties far surpassing those of earlier mediums. Crucially, the digital medium covers the entire communications circuit, including the production (writing, editing) of texts, down to their distribution and even consumption. Moreover, where other mediums can only represent some modalities at best, the digital medium offered the possibility to represent the full spectrum of modalities (text, still and moving images, and sound) in a single medium.

The digital medium's uniqueness is often expressed in terms that compare it to other mediums. Some of the many properties that are often attributed to the digital medium have been mentioned in passing. There is, for example, the extraordinary ease with which endless numbers of copies can be made without loss of quality. But equally, the medium's 'textual instability', and lack of closure—of the typographic form as well as content—are features that set this medium noticeably apart from earlier ones. I would like in this chapter to examine these and other properties in a more structural manner, also analysing exactly where they come from technologically.

It will be helpful to do so by ordering the features that characterise the digital textual medium into a kind of hierarchy. At the top tier of this hierarchical ontology are the inherent or core properties of the computer; below them on the second tier are the technological features of the digital medium made possible by the computer, and at the bottom the social consequences that in turn derive from these technological features.

Core technological properties

The previous chapter described how the Universal Machine was conceived and created in its electronical and digital form, and how this enabled the processing of text—though it took some time for word processing to be developed as such and eventually to find its present widespread acceptance. Also the linking of computers in a network brought a whole series of derivative but very profound technological consequences. These major developments offer the key to the identification of the crucial core technological properties of the computer of which the digital medium makes use. These are (a) that it is a Universal Machine (b) that as a Universal Machine it functions electronically and digitally and (c)

that it operates in a network.

A. Universal Machine

As a Universal Machine the computer is more than just a medium for text transmission, the way for example a book is. As a technology that can manipulate symbols the computer can be used for all tasks for which algorithms can be programmed.[1] That the Universal Machine manipulates symbols is the property of the digital medium that has enabled the convergence of the various medial modalities that exist today. In the same way as text has been made machine readable, this has been done with the other modalities: sound (music, speech), still images, moving images. The computer as a Universal Machine now represents and processes them without significant restrictions in the form of ones and zeros. All modalities are thus fully integrated into one and the same medial environment, for example in a web browser that is equipped with the necessary plug-in modules.

That the computer is a Universal Machine means that it is an instrument capable of performing on text not only all medial functions such as creating, editing, storing, publishing, distributing and consuming it, but also operations of various kinds that never used to be part of Darnton's 'communications circuit'. To take an example, the computer can perform statistical analysis, based on the computation of word frequency and word proximity. In this way, for example, stylistic phenomena, but also authorship issues may be examined.[2] Such 'humanities computing'[3] operations can take place in the same medial environment, either wholly independently of the computer's medial function or, more interestingly, in such a way that it enhances that function, as for instance in the case of a journal presenting access to articles on the basis of tag clouds.

The programmability of the Universal Machine, which is infinite in principle, is also the property that most piques the human imagination. The infinite diversity that follows from that programmability has for example given rise to the expectation (hope to some, fear to others) that the computer may come to rival if not surpass humans in intelligence. In many areas the computer definitely has the upper hand. Calculating is an obvious example in the light of its calculator origin, but it also beats the human brain easily in sheer power of memory. Meanwhile the point has been reached that computers are capable of defeating the best chess players in the world. So far, the computer's growing ascendency over what the human mind can accomplish has mostly been confined to certain well-

---

[1] In her translation of Menabrea's article Lovelace had already said of Babbage's Analytical Engine that this 'may be described as being the material expression of any indefinite function of any degree of generality and complexity' (Menabrea, 'Sketch of the Analytical Engine Invented by Charles Babbage', in *Science and Reform*, p. 267.) On the basis of the program for calculating a series of Bernoulli numbers that she wrote for the never-completed Analytical Engine, Lovelace has often been called the first computer programmer. The claim is dismissed by Anthony Hyman (*Science and Reform*, p. 243).

[2] See for example Hugh Craig, 'Stylistic Analysis and Authorship Studies', in Schreibman *et al.* (eds.), *A Companion to Digital Humanities*, Maldon, Oxford and Carlton, 2004, pp. 273-88. On the basis of such techniques in 1996 the magazine *New York* unveiled the identity of the writer of *Primary Colors*, a *roman à clef* about Bill Clinton's campaign for the U.S. presidency (see Steven Johnson, *Interface Culture*, pp. 152-53).

[3] Part III, 'Applications', of *A Companion to Digital Humanities* (ed. Susan Schreibman *et al.*, pp. 271-468) gives a good overview of different applications of humanities computing.

defined areas. A computer with mental powers that are just as versatile and agile as a human being's is yet to be created. The question of whether computers will be able to possess human intelligence, consciousness and emotions, would not seem to be directly relevant for the subject of this book. Yet it is. All developments in the digital medium occur against the backdrop of the Universal Machine's programmability. When discussing the notion of 'intelligent text' a little later, that metaphor in fact indicates that text on the computer need not stay within the domain of the word processor, but may be the object of a whole arsenal of approaches and processes similar to and beyond the examples just given under the nomer of 'humanities computing'. It is primarily our own imagination that determines the limits of what is possible. The emotional responses that the computer in its guise of machine-as-human has always evoked will be discussed further under the 'Social impact' heading later in this chapter.

B. Digital–electronic nature

The second core property is the digital–electronic nature of the Universal Machine. To create, store and redisplay stored text, and to publish and distribute it, a combination of hardware and software (operating system and application software) is needed. This combination of hardware and software is what I have in Chapter 2 called the 'computing environment'. While the substrate on which the digital text is stored (working memory and storage medium) consists, like the rest of the hardware, of tangible and visible matter, the text itself is virtual in nature. Virtual here means 'existing conditionally'. In more precise, albeit negative, terms, a virtual text is an intangible, invisible and unreadable representation of that text. It is stored in such a way that it may, in certain conditions, be made visible, legible and tangible. One of those conditions is, for example, the availability of electricity.

This virtuality gave text for the first time in history an 'inscrutable' form, like moving images (e.g., video) and sound (a cassette tape) had known for a long time. That is to say that examination of the physical substrate—say, a CD, a memory chip or a hard disk—gives nobarely any impression of the nature or extent of the material that has been recorded on it. This requires specialised hardware and software. That material may be images, music, numbers, spoken word, film, or text, in any combination and quantity. Without the availability of the 'computing environment', consisting of the right combination of hardware and software (and the necessary electricity) the registration might as well not have taken place at all.

In the computing environment text is not only virtual in nature and inscrutable in form, but also machine-readable. To be able to 'calculate' it, text is represented in the computer in the shape of a binary code for each constituent character, space, punctuation mark, etcetera, along with codes for its (typographic) representation on screen or in print. (The computer can also represent text purely as graphics, i.e., as a collection of dots. These pixels are of course also represented in binary form, but the characters they represent cannot be manipulated individually. By applying OCR to the collection of pixels that together represent the text, graphic text can be made machine-readable.) In the era of the

standalone computer there was no standard for text representation; at IBM several character sets were even used side by side in a single company. The desirability of a standard became an absolute requirement when computers had to be able to communicate effectively in a network. For a long time that standard was ASCII (1963); now it is increasingly Unicode (under development since 1991) that has to ensure the interchangeability of texts between different applications and operating platforms.

The property of machine-readability, however obvious, is crucial precisely because that makes it possible to manipulate the text. The digital, machine-readable record thus has the property that it is not permanent. That is to say that a text recorded or 'inscribed' on a medium or in working memory is not actually being recorded or inscribed in the analogue sense, but can be changed or erased at any moment. This is a property that follows from the nature of the computer as a calculating device. If the parts of the text were locked in an unvarying relationship to each other, they would not lend themselves for 'calculations'. It is the fluidity of the virtual data that determines the nature and potential of the digital text. The computing environment itself incidentally always remains available, it never becomes exhausted. Interestingly enough that goes for the text being processed as well. The operations performed on it never exhaust the text, and it can be saved in its original form as well as in any form that results from an operation. Each of the versions preserved can in turn serve as a starting point for further processing, and so endlessly on. The text never gets an unchangeable, final form, except when it is exported out of the computing environment, by being printed out in so-called 'hard copy', for example, or when it is burned onto a CD or DVD (although in that case the text still remains machine-readable and therefore continues to lend itself to further processing; the CD or DVD just makes a 'snapshot' of the text in a given state).

C. Network

The third core characteristic of the digital medium is that it exploits the possibility of communications between computers. Strictly speaking, this is of course not a core property. After all network communication is one of the infinite number of possible applications of the computer as Universal Machine. The main reason to regard this particular application as a core feature nevertheless, is a pragmatic one: namely that it is at the basis of so many other important properties. This emphasis on the network does justice to the enormous significance of the communications dimension of the digital medium—in general, but especially for textual transmission.

As recounted in Chapter 4, the Internet was created in 1969 when the first two computers of the ARPANET, designed by the Advanced Research Projects Agency of the US Department of Defense were linked to each other. Soon flow-controlled protocols for two-way traffic became a networking standard. This means that the data to be sent are divided into packets of a certain size, which may be sent via any route to the specified address. The protocol keeps a continuous tab on the dispatch and safe arrival of the packets between the sending and receiving computer, and rejoins the packets again after arrival. Since 1978 the TCP / IP set of protocols has been used for this purpose. Although

developed for military purposes, the network was soon used for scientific communication of all kinds. In 1991 started the publicly accessible World Wide Web invented by Tim Berners-Lee. In the context of the communications circuit, the special significance of the linking of individual computers in a network is that this network lends itself not only for the creation and production but also for the *distribution* of machine-readable, manipulable, virtual text.

The 'Docuverse': the information space of the Internet as a medium

The combination of these three core properties in the digital medium has led to a whole new kind of information space. This phenomenon is well covered by the term 'docuverse', which originates from Ted Nelson's concept of hypertext.[4] In order to delineate the contours of the docuverse I will try to characterise that space based on a number of second-tier characteristics derived from the technological core properties described above. Drawing parallels with the press (and other mediums) will bring the implications of these core technological properties into sharper focus.

Speaking in terms of advantages and disadvantages always carries the danger of limiting one's perspective, since they are inevitably advantages and disadvantages compared to a standard, which often remains implicit. Moreover, such a standard is usually not fixed, because humans are inclined to view advantages and disadvantages mainly in the light of their present circumstances. Also advantages often have an unpleasant and not very predictable tendency to turn into disadvantages in the longer term. (The opposite happens as well, but unfortunately that seems less often the case.) Nevertheless, the use of the terms make sense. As Spinoza already makes clear in his *Ethics*, humans can only judge their interaction with the world around them in terms of what is advantageous versus what is disadvantageous or causes hurt—a form of Darwinian natural selection. In this characterisation of the Internet as a medium I will inevitably use the perspective of the *homo typographicus* that someone of my generation is and remains. But in so doing in any case I have named my standard. I will try to apply it as openly as possible and without personal prejudice. The question whether the properties I will be discussing are intentional or unintentional is as relevant here as it was in the cases of writing and printing discussed in Chapter 3.

In a comparison with the communication circuit of the printing press a large difference in functionality becomes obvious. The first of the three core properties of the digital medium is that, being a Universal Machine, the computer—the hardware heart of the computing environment—is infinitely functionally expandable. The printing press on the other hand does only one thing: it multiplies. In the constellation of operations which lead from the creation stage to the finished product many other technologies besides the printing press play a role. For the writing, distribution and consumption of text the printing press is not suited. Also, a printing press can only handle text and still images, while the digital medium can process all modalities of all other mediums. Just as the

[4] Ted Nelson, *Literary Machines*, [the author], [original edn 1981] 1993, pp. [4/15].

limitation of the printing press was unintentional, the computer became the unlimited and comprehensive medium that it is now in a fairly random way.

The new textual instability That the computing environment is of a digital–electronic nature I have previously defined as the core property responsible for the virtuality, inscrutableness, machine-readability and manipulability of text (and other data). Here the instability of the digital text stands out as a crucial difference compared to the products of the printing press. The printing press has in the course of time created a (largely unconscious) expectation of stability and permanence of form and content. (That this took time, and that stability, but also reliability, were not self-evident in the early days of the printing press, as Adrian Johns has convincingly demonstrated, does not alter the outcome of the historical process.[5] This outcome was suggested—if not predetermined—by the salient properties of the printing press.) The virtual nature of the digital text, however, works against such closure. The digital text can keep changing shape constantly. Form, content, and even the existential state of digital text may at any time and with the slightest effort be changed. What often tends to be regarded as one of the greatest achievements of print thus falls away: the stability and permanence of the textual foundation on which we have raised the cathedral of our culture, and science and scholarship in particular. All new knowledge that humans acquire builds in part on the corroboration or refutation of existing knowledge. That that previous knowledge is fixed and that it can always referred to is the security society has learned to trust as the world of manuscript and Bible was gradually overtaken by the Order of Books and scientific rationality.

The lack of closure—compared with print—of the digital text is dual faced. Compared with the modern book publishing process the information space of the Internet as another medium in which texts are published and distributed is extremely unstable. Machine-readability brings with it boundless manipulability, which can be extremely difficult for *homo typographicus* to cope with. At the same time, this manipulability is a very useful function of machine-readability. Word processing is based on it, but beyond word processing, the potential of the lack of closure and the lack of permanence of the digital-text is huge. It is due to the influence of our typographic history that it is only relatively slowly being recognised and mobilised as a valuable property. To the potency of the lack of closure I will return under the heading of 'intelligent text' below.

The end of the copy From the description of the virtual nature of the digital text it was already implicitly evident that it follows from that virtuality that unlimited copies can be made of the text without deterioration and without significant cost. This property was also not planned, and this time, too, it represents a Janus head. The architecture of the Internet ensures that transmission of data in fact creates a clone of the original data on the receiving machine. To be able to read a Web page the browser on your computer (the client) makes a copy of the data residing on the server. Sending a file from one computer to

[5] In *The Nature of the Book* Johns has stressed the social process through which the expectation of stability and reliability was able to evolve. He elaborates on the instability and confusion that characterised the early period of printing. Johns suggests that the exceptions to the stability that the printing technology brought with it, for example due to the phenomenon of correction on the press and piracy, were actually numerous enough to make instability to the rule (see, e.g., p. 31).

another, for example via email or ftp, ensures that after the transmission is completed the receiving computer contains an identical copy of the original file. However often a document or page is requested by users, this never exhausts the document. In analogue terms, the digital document has as it were a built-in copying press, which manufactures a copy for any potential reader. Not only are these copies made at extremely low production and distribution costs, but they are in fact so perfect that the fundamental distinction between original and copy is no longer relevant. The advantages are obvious. The disadvantage—in the terms in which we are now accustomed to think—of the extreme ease with which perfect digital copies can be made and distributed in the digital information space, is chiefly the threat it poses to copyright.[6]

As early as the 1970s the traditional concept of authorship was denounced by critics like Michel Foucault, Roland Barthes, and Jacques Derrida. The ease of digital copying, cutting and pasting came just at the right time to illustrate in daily practice what they had already claimed in a philosophical and theoretical sense. If the author was not already dead in theory, it was time for a fundamental recalibration of his status for reasons of actual practice.[7]

Digital distribution As a stand-alone word processor, and even as a layout and typesetting machine, the computer was never more than a tool in the existing print production process. The use of word processors and lay-out programs led to greater efficiency, but the entire process remained focused on eventual reproduction, through the printing press or laserprinter. It was the communication between computers in a network that gave the big push to the emergence of a new medium in which distribution (and consumption) of machine-readable virtual text in a digital form was possible. The digital medium thus seamlessly integrates all functions from Darnton's communications circuit in a single environment. Moreover, digital distribution takes place without significant cost to the parties directly involved (see Chapter 4).

Architectural flatness That the network architecture of the burgeoning Internet addressed strategic military applications and specifications from the beginning, was of crucial importance. To ensure that in case of damage to the network the data traffic could always continue the Internet is non-hierarchical and flat. Interestingly, this architecture also became closely associated with the egalitarian hacker culture of the computer pioneers.[8] In an environment where all other users were in fact by definition peers, the only restriction in the use of the network was the technical knowledge required to access the Internet as a medium for sharing and making public text. As more people gained access to computers and the Internet, that knowledge became more commonplace. Soon software was written for a variety of functions that placed the network at the convenient

---

[6] Dirk Visser, professor of intellectual property law at Leiden, even suggested in his doctoral dissertation *Auteursrecht op toegang: De exploitatierechten van de auteur in het tijdperk van digitale informatie en netwerkcommunicatie* (Intellectual property right on access: The exploitation of the author in the age of digital information and network communication; Den Haag, 1997) to remove the ban on reproduction from copyright law.

[7] It is ironic in this light that, as we have seen, the digital information space precisely yielded efficient new means to solve questions of authorship.

[8] Castells, *The Internet Galaxy,* p. 14.

service of all and sundry for the creation, production, distribution and consumption of text. Thus, things were steadily made easier for the user. When Tim Berners-Lee in 1991 launched the World Wide Web, it presented few barriers to prospective users. Since then, so many online publication tools have become available that no technical expertise is required at all.

Two-way traffic The architecture of the Internet is based on two-way traffic between a client and server computer. But in principle, any computer can be both the client and server (and, being a Universal Machine, have an unlimited number of other functions). The existence of this continuous contact between client and server—which can be logged—constitutes an important difference to the situation in the world of physical texts. Publishers who distribute books through bookshops have no idea where these books end up. The process is one of essentially one-way traffic. I will return to the subject of logging and its uses in 'The Docuverse and the Universal Machine' below.

Low cost Besides ease of access the costs of Internet use are also low. Governments and scientific institutions have invested heavily in infrastructure, a form of cost allocation which was partly motivated by the concept of the digital highway as a parallel with physical roads. Together with the dropping cost of disk space and the fact that little investment in production is required, this makes the World Wide Web as a publishing medium particularly cheap. Again, the comparison with print is instructive. The nature of print technology makes the costs involved in its production and distribution high. The technical knowledge needed for printing, for example, has always remained relatively high, and the requisite training is expensive. Some of those costs have also to be made digitally (e.g., editing and design), but a significant portion of the costs is mostly or even entirely absent: the initial outlay in reproduction by means of printing (paper, printing costs, binding) and the cost of physical distribution. The low cost of publication on the Internet leads to a low economic threshold, and thus to a rapidly increasing amount of published information. However, some infrastructural constraints remain, especially in poorer countries, but also in less densely populated areas.

Speed The transmission reach of the network is now global and the transmission speed of broadband so high that any computer connected to the network anywhere in the world can be reached almost immediately.[9] The place where information is consumed may be different than the one where the publication takes place—and that in turn can be different than where the creation took place. In this way the digital information space has, in fact, made both distance and time irrelevant. The combination of range and speed is now making place-shifting possible, in addition to the already existing concept of time-shifting introduced by the VCR.

Although the many-to-many architecture is difficult to compare to the traditional

[9] Internet usage, population statistics and Internet market research data for over two hundred individual countries and world regions may be found at http://www.internetworldstats.com. The broadband statistics (http://www.internetworldstats.com/dsl.htm) show that in some countries up to one third of Internet users had broadband connections in 2007.

one-to-many architecture of the broadcast mediums, the new medium with its wide distribution and high speed has broadcast potential. That means that in principle it is possible that many users simultaneously 'tune in' to the same material, held on one or more other websites. By means of so-called 'streaming media' successful broadcasting experiments have already taken place. In addition, the Internet is increasingly used to make archived radio and television broadcasts available as streaming media or downloadable files.

Convergence In the digital medium all modalities converge. With the emergence of digital distribution via broadband the convergence of all vertical medium columns (music, radio, television, newspapers and magazines, books, and now also games) could, in principle, take place. But that the technological possibility of such far-reaching convergence exists, does not automatically entail social acceptance. Whether, when and to what extent convergence will actually happen, remains to be seen. For example, it is possible that consumers will prove to have a strong preference for combinations of certain (dedicated) devices with certain types of information (or modalities). Despite the potential of various devices to integrate a variety of functions, many people for example appear to prefer to use devices tailored to a single specific function (such as phone, music player, PDA) rather than one device with multiple functions. In addition, analogue mediums will not disappear overnight. The fact that fiction, for example, is still consumed in printed form much more readily than scientific journals or reference works may not quickly change, in which case its production and dissemination are likely to remain predominantly analogue.[10]

Access through content Through the possibility of querying the full-text contents of all documents in the digital information space, access to information is being transformed. Using keywords and phrases as search terms gives direct access to passages on Web pages, but also in digitised books, bypassing the need to go first through traditional bibliographic methods and then to locate the relevant passage by reading the whole text. In this way the digital information space not only provides access to entire books and articles, in the way a library catalogue does, but also directly to passages within these books and articles. (This interestingly applies also for books sitting on a bookshelf at home which, given the same edition, can be accessed in the same way as their digitised counterparts, by searching them full-text on the Internet.)

Besides their storage and preservation function providing access to information has always been a crucial function of libraries in the value-chain. Filtering, organising, and creating metadata are the main grounds for libraries' existence. They carry out these tasks with the help of the sophisticated system of bibliographies and library catalogues made familiar by the Order of the Book. Searching and finding on the Internet on the other hand completely bypasses traditional ways to gain access to texts through the bibliographic apparatus and so also ignores aspects of validation and certification of the content. Compared with an actively filtered and orderly information environment such as the

[10] But note that digital printing and 'print-on-demand' can make use of distribution via the Internet.

library the information space of the Internet can be regarded as a kind of black box, where the most reliable information is thoroughly mixed up with materials of more doubtful status, without context and without any indication of their origin or trustworthiness. The extra demands placed on users by their active role in the discovery and assessment of such information whose provenance and integrity remains unclear will be discussed more extensively later.

The possibility to locate text in the digital information space through full-text searching (whether in a tailored environment such as Google Books or through a full-text search on the Internet at large) can be said to signify the end of 'the document'. Documents that were originally physically separately published now form a de facto unit in the digital information space or 'docuverse'. The continuity of this textual space represents a fundamental difference in relation to the world of material documents, where physical separation also means logical separation. If the Library of Alexandria was to function as the information space it was designed to be, its books had to be gathered in the physical location of the library. The Internet brings together virtually, in any desired location, information that may physically reside in the most diverse places in the world.

The docuverse and the Universal Machine

The fact that the 'communications circuit', familiar from the analogue world, can also be used to model digital communications suggests a high degree of continuity between the world of books and the docuverse. But there are discontinuities, too, for example in the way the docuverse lends itself to searching and finding *inside* texts. However, the chief cause of discontinuity is probably that the docuverse is not just an information space, but is also only one of the digital equivalents of activities and operations from the communications circuit that have been made possible by the Universal Machine. Not only the search methods that the digital environment offers are much more advanced than the older analogue ones. The same goes for many other digital translations of analogue practices. These are advantageous in that they are usually faster, easier, more convenient, more exact, or otherwise 'better' ways to reach the same or similar results. But apart from such more or less equivalent activities and operations, the Universal Machine also offers possibilities for the treatment of text (and other modalities) that were not available in the traditional communications circuit at all. Furthermore, these advanced additional possibilities are available in the same docuverse where the digital communications circuit also takes place. Under the general heading of we might be called 'the creation of knowledge about text', I want to discuss a couple of categories of ways to create and store that knowledge.

In Chapter 2 under the heading 'Markup' text encoding was presented as an alternative to typography to indicate the structure of text. That is how, for example, much of the HTML coding of web pages works. But more sophisticated markup languages, based on XML, are capable of much more. Markup languages can 'encode any relevant information about the text in such a way that information can be understood and processed by a computer', as Chapter 2 expressed it. The application of analytical

information about text through markup can be succinctly referred to with the phrase ‘making text intelligent’ or ‘making a text knowledgeable about itself’. This is a shorthand way of saying that markup can serve to instruct computers to recognise certain constituent parts of texts, to teach them the meaning of parts of texts, and to make them do things with that ‘knowledge’.

As Chapter 2 made clear, the type of information about the text that can be captured ranges widely, and includes information about the structure of the text; about its typographic design; about its interpretation, as expressed in editorial notes, standardisations, glosses; and so on. By far the most important of these is information about the structure of the text. That is to say, information about the exact function and mutual relationships of all the elements of which the text is made up. The emphasis on structure can be well illustrated by the case of a doctor’s prescription. Here the unambiguous encoding of the structure of the document ensures that, for example, the patient data are not confused with information about the doctor or the pharmacist. The contrast between, for example, handwritten and printed information that ensures in the paper world that the data on doctor and patient are not mixed up, is replaced by digital markup codes which makes explicit the structural function of each constituent part.

While in this example form is almost irrelevant, in most day-to-day circumstances (typographic) form is hugely important. Markup in such cases also serves as an alternative to the function of typography in analogue text. Information about the structure of the text offers the possibility, for example, to make the presentation of the text independent of output and device. Once encoded in XML a text is medium independent and by using a style sheet it can for instance be converted to a typographical format such as PDF (for printing or screen), but also to another markup format, such as HTML (for the Web) or EPUB (for an e-book reader). The design simply adapts to the requirements of the device. The design of a whole series of texts can be adjusted by changing a single style sheet.

The degree of objectivity with which the structure of documents can be captured in markup depends on the level of depth required, as well as on the nature of the text. The information contained in a doctor’s prescription will leave little room for subjective interpretation. The same applies, in line with their already pre-structured nature, to forms. In other texts there will be more room for interpretation. A novel, for example, would tend up to a certain level to have a clear structure: there are preliminary pages created under the publisher’s responsibility, containing, for example, a title page and publication data; the main body of the work supplied by the author, for example divided into chapters, and maybe some end matter. But within the chapters of the main work there may be a structure, too: a narrative one, for example. Themes, persons, acts and events may form part of the structure. The analysis requires interpretation, which is by its nature subjective. On the whole, the older the text, the greater the role of interpretation. This is the arena of textual scholarship, which also gratefully and intensively makes use of the analytical capabilities of descriptive encoding.

The intellectual effort needed to apply editorial notes, standardisations, glosses and so on, is not fundamentally different in nature in a digital than in an analogue environment. Regardless of whether the results are displayed by means of analogue footnotes or digital markup, the same text analysis and interpretation have to be

performed. Where there is a substantial difference is in the wider possibilities that markup offers for the computer processing of interpretations that have been added to the text. For example, an editorial apparatus may or may not be included in a text edition, or only those notes may be displayed which are relevant in a particular context; or an index of relevant terms or names may be automatically generated.

Though the processing possibilities are greatly broadened by the computer, the concept of the creation and publication of an authoritative text enriched with interpretive expert knowledge by an appropriately skilled person is thus not new. In that sense, the practice of what I have called 'making text intelligent' is still strongly motivated by a notion of scholarship that was fully determined by the Order of the Book. But the Universal Machine offers vistas of entirely different ways of dealing with text and textual meaning. That brings us to the edge of the communications circuit, familiar from the world of the book, to enter the field of humanities computing—even if there are, as previously noted, all kinds of links.

Actively making text intelligent has so far continued to rely on individual expert interpretation and is therefore labour intensive. Instead of making text intelligent, it is also possible to apply artificial intelligence to text. That is to say that the computer can be instructed to analyse text on our behalf. An example is the commercial application of statistical analysis by Amazon:

> Capitalized Phrases, or 'CAPs', are people, places, events, or important topics mentioned frequently in a book. Statistically Improbable Phrases, or 'SIPs', are the most distinctive phrases in a book. Just as CAPs and SIPs give you a quick glimpse into a book's contents, a movie's actors, directors, and plot keywords give you more information about that movie.[11]

These kinds of uses of the computer are perhaps a better exponent of the new digital order than making text intelligent by hand. The main challenge to the use of forms of artificial intelligence is building the instruments. Once they have been built, they can be applied to any text present in the docuverse. Thus, the results can immediately increase exponentially.

A third way in which the Universal Machine can learn about text and then make that information available to humans shifts responsibility for supplying that information from a limited group of experts to web users in general. The principle of the wiki is a well-known example of the 'democratic' way of adding knowledge to the web.
In the case of Wikis this concerns new knowledge. But according to a similar process also information about existing texts could be added. For once texts are available digitally, they can always be further processed and edited. The digital information space is ideally suited to break the hierarchical relationship between an author or editor and the reader—which is also still at the foundation of the 'conventional' process of making text intelligent. Instead of delivering himself up to the passive consumption of an interpretation generated by an expert author or editor commissioned by a publisher, the consumer can exercise similar

[11] Amazon website, http://www.amazon.com/gp/phrase/help/help.html.

activity himself. Different people can comment on same digital text, giving rise to, for example, various—virtual—combinations of texts and commentaries. Not surprisingly, in view of the long-established and familiar hierarchical system of knowledge generation, this is not being widely adopted yet.

There is a fourth way to let the Universal Machine aid in textual interpretation, and this is potentially the most powerful one. This involves tapping Internet users' online activities, which may be totally unconnected with the textual interpretation question at hand. A good example is the way Google uses so-called CAPTCHAs to improve optical character recognition (OCR) in the Google Books programme. CAPTCHA stands for 'Completely Automated Public Turing test to tell Computers and Humans Apart', and it is familiar to Internet users from the images of deformed characters that have to be deciphered so that online companies know that they are dealing with a real person. Google works with a company called ReCAPTCHA,[12] which takes its word images from scanned print materials. Google can compare the result of deciphering an image whose text it already knows with an image whose text it doesn't know, and so the byproduct of the deciphering act is the solution of an OCR question.

In these various ways, the Universal Machine can 'learn' about the content of texts and place its 'knowledge' in the hands of humans. Thus slowly the contours can be seen to emerge of what Tim Berners-Lee has called the semantic web. This is a web where information can be analyzed and interpreted entirely by computers. By combining the highly flexible and analytical XML with description techniques such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL) conventional texts can be transformed into machine-readable descriptions of the data that these texts contain. In so doing they represent the knowledge that humans have about those data, but because those descriptions are machine-readable, they offer powerful ways to process the knowledge contained in those texts with the help of the computer. This also opens up prospects to instruct the computer, for example to apply formal logic to these texts. Although daily practice does not yet make full use of the possibilities, the technologies to achieve this kind of scenario exist now.

Another dimension of the use of the Universal Machine in the docuverse is that not only the content of texts, but the way they are used can be analyzed. The logging of the user traffic in the digital information space is again dual-faced. That is to say, there are both very valuable and very dubious reasons for logging the use of digital sources. One useful purpose, for example, is to improve the functionality and the nature of materials on offer. The owner or curator of a digital collection can adapt the presentation of materials from the collection to the implicit user wishes that emerge from an analysis of the user traffic. There are obvious drawbacks associated with this functionality. As all web users know from experience, web use always leaves many traces behind. Sometimes users are aware that their use is being logged, for example when they instruct their internet browser to accept cookies, small text files that are stored on the hard disk of a client visiting a website, which store information about the visitor's activities. But even without such an intimate relationship between server and client, visits and visitor behaviour can be

[12] ReCAPTCHA is a company cofounded by Luis von Ahn, about whose work more will be said below.

recorded. Web statistics can be maintained using IP addresses. A very basic example is the fact that many Web sites (like Google) adjust their ads to the client's geographic location. In most cases the user will not be aware which web activities are being logged and by whom. The semantic web, too, may lead to greater erosion of privacy, and more readily accommodate attempts at censorship. Armed with text-analysis software, authorities of all kinds may, legally or illegally, make good use of advanced semantic web description techniques to scan web pages in an automated fashion for content deemed offensive, dangerous or unacceptable in any other way.

In the hands of a harsh government logging the user becomes a kind of invisible digital counterpart of Jeremy Bentham's panopticum. The efficiency of the docuverse in this regard is such that Big Brother can meet all its surveillance needs with very little effort. Criminals, too, can benefit from users' generally rather limited awareness of the number of tracks they leave behind. And even if that is not problematic at the moment itself, it may become so at some later time. In the docuverse a forgotten activity or unguarded remark may surface at any time. Because it is almost impossible for users to cover their tracks, this can all too easily lead to a loss of privacy. (To many people this is not an issue; see 'Private domain becomes public domain' below.)

Social characteristics of the docuverse

Loss of privacy is one of the more recently recognised insidious social consequences of the digitisation of our everyday existence, but there are countless other consequences, both more and less desirable ones. Not surprisingly, the advent of the Internet and the growth of the World Wide Web as a medium has induced and continues to induce a neverending stream of social commentary. The awareness to what extent humans are the product of their history is greater than ever before and the role of mediums in that history has been thoroughly investigated. Especially our dramatic experiences with the demagogic force of mediums in the course of the twentieth century have given us an extra sensitive medium antenna. However, the main focus of the attention has been on the way the 'mass media' have been used: much less on the intrinsic properties of mediums, and even less on the intrinsic properties of the *textual* mediums—with the (partial) exception of newspapers. In the early days of the Internet text was the preponderant modality, and little 'mass' was involved. The Internet was not a public medium. It came from the strategic interests of the U.S. Department of Defense and the scientific world. After the rise of such mass mediums as film, radio, and television, the demagogic potential of the Internet did not initially seem particularly large. Yet from the start people assigned to this new technology farreaching social implications. It was soon predicted—correctly, as it turned out—that its impact on our culture would be no less radical than that of Gutenberg's movable type. Though this was not in itself an outrageous prophesy, it was nevertheless possible to make exaggerated claims for the new medium. For example, Ted Nelson characterised the subject of his book *Literary Machines* in this way: 'This book describes the legendary and daring Project Xanadu, an initiative toward an instantaneous electronic literature; the most audacious and specific plan for knowledge, freedom and a better world yet to come out of

computerdom'.[13] The high expectations concerned especially the revolutionary possibilities that the Internet and the hypertextual precursors of the World Wide Web offered for interactivity and collaboration. Instead of passive consumption the digital medium promised to promote active participation in the textual discourse. The medium would bring about a flourishing 'wreader'-ship.

It took some time before the World Wide Web, although it was built on the very foundation of hypertextual linking of information, began to honour its promise of hypertext for the masses. The use of the HyperText Markup Language (HTML) made the World Wide Web indeed hypertextual. Initially, however, its actual functionality remained far behind what hypertext theorists like Ted Nelson and George Landow had conceived. There was hardly any two-way traffic and collaboration was only indirectly possible. Indeed, the average Internet user took very little interest in the opportunities for collaboration that lay behind Ted Nelson's original idea of hypertext. It therefore took some time before the opportunities for cooperation offered by the Internet with its flat architecture, democratic access and two-way traffic with global reach were also used outside the scientific community.

An important factor in this delay was the rapid colonisation of the Internet by commercial interests. Commercial companies approached the Web mainly from a traditional industrial top-down perspective (from producer to consumer, from seller to buyer). In that model interactivity had no obvious place. Only gradually realisation grew that bi-directionality was a salient technological property of the medium, with a dynamics of its own. Since then, the two-way traffic is being used in a more creative way, especially by the largest and most successful 'e-tailers', such as Amazon.com or Apple, and the possibilities for interactivity have been improved. Instant messaging and blogs are very popular, while wikis offer an interactive writing and editing environment for just about every conceivable kind of use. When the possibilities of interactivity began to crystallise and the democratic potential of the Internet was recognised more broadly, commercial companies chimed in by devising new business models and scenarios which were less top-down oriented and were better suited to the nature of the technology.[14]

But the advent of the Internet inspired not only optimistic scenarios. Its emergence as a medium invoked many more sombre reactions. One of the more eloquent and better-known ones was Sven Birkerts' early pamphlet, *The Gutenberg Elegies: The Fate of Reading in an Electronic Age* from 1994. Invoking an ominous quotation from Antonio Gramsci, Birkerts situated culture in these digitising times in a place, wedged between books and the digital medium, where 'the old is dying and the new cannot be born'.[15] Besides a certain disdain Birkerts' book notably expresses a grave concern that the computer as a technology could well have all sorts of undesirable consequences:

---

[13] Cover of the 1993 edition. See also George P. Landow, *Hypertext: The Convergence of Contemporary Critical Theory and Technology*, Baltimore and London, 1992, p. 19.

[14] That the main business model is shifting from paying for content and intellectual property to other models, including paying for advertising or 'freemium' (see Chris Anderson, *Free*, New York, 2009) is a source of concern about the erosion of a conventional professional information environment. *The Cult of the Amateur: How Today's Internet Is Killing Our Culture* (London and Boston, 2007) by Andrew Keen is meant as a warning for the harmful consequences. I will return to this later in this chapter.

[15] Sven Birkerts, *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*, New York, 1994, p. 121.

> A change is upon us—nothing could be clearer. The printed word is part of a vestigial order that we are moving away from—by choice and by societal compulsion… This shift is happening throughout our culture, away from the patterns and habits of the printed page and toward a new world distinguished by its reliance on electronic communications… The evidence of the change is all around us, though possibly in the manner of the forest that we cannot see for the trees. The electronic media, while conspicuous in gadgetry, are very nearly invisible in their functioning. They have slipped deeply and irrevocably into our midst, creating sluices and circulating through them. I'm not referring to any one product or function in isolation, such as television or fax machines or the networks that make them possible. I mean the interdependent totality that has arisen from the conjoining of parts-the disk drives hooked to modems, transmissions linked to technologies of reception, recording, duplication, and storage. Numbers and codes and frequencies. Buttons and signals. And this is no longer 'the future', except for the poor or the self-consciously atavistic—it is now. Next to the new technologies, the scheme of things represented by print and the snailpaced linearity of the reading act looks stodgy and dull. Many educators say that our students are less and less able to read, or analyze, or write with clarity and purpose. Who can blame the students? Everything they meet with in the world around them gives the signal: That was then, and electronic communications are now.[16]

The stealth by which Birkerts perceives the electronic medium to be invading society clearly adds to his fear. It is that same invisibility already identified in Chapter 1 as one of the problems that especially the textual mediums face, but appearing now in a more sinister guise.

A similar fear spoke from Neil Postman's *Technopoly*. Postman believes that in our 'technopolis' we allow ourselves to be dominated by technology more generally, of which computers are merely the most advanced and threatening example. Though Postman's book is about technology in the broadest sense, textual mediums (manuscript, print) and language receive his special attention because, as already suggested in Chapter 1, they play such a decisive role in how we interpret and experience the world, precisely because of their invisibility. Maybe Postman's reservations are primarily motivated by the fact that word processing, and in even stronger measure the Internet, engage the computer deeply in the sensitive field of human communication.

Fear of technology is of all times and the arrival of each new medium is accompanied by cultural pessimism. Plato's objections to the artificiality of writing, which was also a technology after all, bear witness to this, and the art of printing elicited similarly gloomy responses. But techno-phobia disappears and the criticism is silenced once a certain familiarity grows. Also, the nameless fear of the computer was probably greater while it still stood at the exclusive service of scientists in white dust coats, invisible to the rest of the world. The plot of *2001: A Space Odyssey*, for example, was driven by the scary vision of HAL, an artificial intelligence with all too human features, trying to impose its—or *his*—will on humans. However, this fear has still not altogether disappeared. That has

[16] Birkerts, *The Gutenberg Elegies*, pp. 118-19.

not only to do with the fact that the computer is still relatively new. The technology is also, unlike writing and printing press, not static. Based as it is on the Universal Machine, it continues to evolve, and so continues to be capable of springing new—and potentially unpleasant—surprises.

The Universal Machine as automaton

One of Walter Ong's central propositions is that writing, too, just like the printing press or the computer, is a technology, a tool of *homo faber*.[17] That writing was the first step in the 'technologisation of the word' is a useful insight that has found a wide reception in recent decades. In this book, too, I have stressed the continuity from writing through printing to the digital processing of text. But writing, printing and digital text should not be uncritically equated as so many forms of technology. There are essential differences between these technologies. The disjunction between writing and print, and again between print and the computer, is greater than the concept of the technologisation of the word—with its implication of underlying continuity—may suggest.

However important the role of the printing press has been in cultural history, however many spirits it may have helped to release from their bottles, it has never been more—*pace* Elizabeth Eisenstein—than a passive instrument in the hands of man. This is where the comparison of digital text with writing and printing as technologies is flawed. In the first place the computer as Universal Machine has many more capabilities. Not only more than the printing press, but *more than any other human invention* to date. Its flexibility alone makes it into an unimaginably powerful tool. The computer as a technology harbours a fundamentally infinite number of as yet unimagined capabilities. Secondly, the computer's sophistication is already such that it may be called an emergent life form.

The inscrutability of the computer offers a direct parallel with the human mind: 'It has invited speculation on a special relationship between computers and the equally inscrutable brain.'[18] Humans have an irresistible tendency to antropomorphise, which is stronger as the object in question has more human features. Involuntarily we assign human characteristics to computers, especially when they perform tasks in the field of language and communication. That is not limited to the way in which one finds oneself talking peevishly to one's computer when it dictatorially imposes its blindly mechanical way of working. We are faced with the inescapable fact that the computer as Universal Machine is able in principle to surpass humans in intelligence. Maybe a machine like HAL, with its eerily human features, its 'insecurity' and the manipulative behaviour with which it is 'deliberately' trying to deceive the crew of the spaceship is not (yet) realistic. Nevertheless, the combination of artificial intelligence and robotics has already led to a large arsenal of machines that show aspects of human behaviour. The distrust of critics like Postman and Birkerts is not incomprehensible, and not unfounded.

During the sixteenth, seventeenth and eighteenth centuries the world witnessed

[17] Walter J. Ong, *Orality and Literacy*, pp. 81-83.
[18] Sherry Turkle, *The Second Self: Computers and the Human Spirit*, New York, 1984, p. 22.

what E.J. Dijksterhuis has so appositely called the 'mechanisation of our world picture'.[19] In that mechanisation the printing press played a major role. Its role began early in the sixteenth century, with the widespread dissemination of the Lutheran heresies, and grew steadily, with the appearance in print of ever more scientific and philosophical reflections that equally gnawed at the authority of the only true church. They undermined the divine order, and not only our world picture but also our picture of ourselves as human beings in that world came under heavy fire. With his notion of the dualism of body and mind Descartes had tried to safeguard some of the special status of humans as the highest aim of God's creation. Although the body had to obey the same mechanisms as the rest of the world, the human mind, through its immortality, participated in the divine. Spinoza made short shrift with this makeshift solution in his *Ethics* by stating that spirit (thinking substance) and body (extended substance) obeyed the same laws, and were merely two different manifestations of a single substance:

> Here, before we proceed any further, we must recall what we showed above: namely, that whatever can be perceived by an infinite intellect as constituting the essence of substance belongs to a unique substance alone, and consequently that thinking substance and extended substance is one and the same substance, which is understood now under this and now under that attribute.[20]

A century later Julien Offray de la Mettrie carried Spinoza's insight into the materialistic nature of the human mind to its logical extreme, concluding that people were nothing more than machines.[21] That people were revealed to be mere mechanisms left nothing of the divine that they had always perceived in themselves. The human-as-machine notion—the microcosmic counterpart of the macrocosmic discovery that the earth was not the centre of the universe—definitively robbed human beings of their immortal soul and of their place at the head of creation.[22]

Barely did we get over the shock that the human-as-machine is no more than a mechanism, or we are about to be overtaken by the machine-as-human. It is not easy for humans to accept their limitations. But there is one crucial difference between the realisation that people are essentially nothing more than a machine and the idea that a machine might just become 'human'. In the case of the human-as-machine notion the point was the discovery and then acceptance of a pre-existing reality. The irony of the development of the machine-as-human, however, is that this is entirely of our own doing. In making this invention we might be said to have hoisted ourselves with our own petard. Instead of the failed humanoid Frankenstein, we are now well on the way to build a superior intelligence. Ultimately, according to futurists like Ray Kurzweil, it will be

---

[19] E.J. Dijksterhuis, *The Mechanisation of the World Picture*, OUP, 1961.
[20] Spinoza, *Ethics*, Part 2, Proposition 7, Scholium (quoted in the translation of G.H.R. Parkinson, OUP, 2000, p. 118). See also Part 3, Proposition 2, Scholium.
[21] In *L'Homme machine* (1747), translated as *Machine Man* in *Machine Man and Other Writings* by Ann Thomson (CUP, 1996).
[22] Still today new evidence continuously comes to light that puts the special place of humans in creation in doubt. That chimpanzees are indeed 'prepared' for speech is just one of the most recent examples (see Taglialatela, *op. cit.*). In *I Am a Strange Loop* Douglas Hofstadter has made similar observations about the sliding scale on which awareness can be attributed to humans and other living creatures.

possible with our technologies to 'improve' human beings and achieve 'trans-humanism'. Our situation, however, is more perilous than that of the sorcerer's apprentice, because this time there is no sorcerer who can help us to get back control over runaway technology. The computer is not the only example; with genetic engineering and nanotechnology we have maneuvered ourselves into the same predicament.

As early as 1984 Sherry Turkle observed that '[c]omputers are not good or bad; they are powerful'.[23] Of course that is so, but they *do* possess certain technological characteristics, intended but also unintended, which stimulate certain developments more than others. Our problem—our tragedy—is that we do not know all the properties before they actually manifest themselves, and so cannot predict the developments through which they will carry us. Our relationship to technology is ambiguous. On the one hand with the progression of our technological ability we take fate increasingly into our own hands. This fosters a feeling of power and control. But ironically, that control also confronts us all the more with the fact that we are ultimately responsible for all that we make of ourselves and all that we do.[24] This is all the more acutely painful because we are so imperfect as creators. Again and again we seem to lose control over things, things do not work as we had imagined, or they have unexpected side effects.[25] In this regard it is instructive to reflect that the prediction that the computer would bring about the paperless office has still not come true. (Nor does it appear that the invention of teleconferencing will lead to less travel; rather the opposite will prove the case. The greater our virtual world becomes, the smaller the real one will seem.)[26]

Social changes in the digital textual space

It requires little imagination to see that whether intended or unintended, the practice of reading and writing, and our whole concept of literacy, are strongly influenced by the digital technology. Musings on technology-as-threat represent a fascinating, valid, and possibly healthy because sobering wider perspective on the relationship between technology and social change. However, since they concern only the computer's first-tier properties, i.e., those of the Universal Machine functioning electronically and digitally in a network, they lack specificity. All kinds of phenomena are making their appearance in close conjunction with digital technology. These include the zapping attitude to text, the sheer amount of immediately accessible documents, mostly hypertextually linked, that are already present in the digital information space; the rapid increase in the volume of that information space; its uneven quality, and the instability of digital textuality. In the remainder of this chapter I intend to examine these social phenomena somewhat more systematically, and in greater detail, regarding them as social consequences of the second-

[23] Turkle, *Second Self*, p. 323.
[24] Alain de Botton has expressed this quandary very well in his *Status Anxiety* (London, 2004); see also Jos de Mul, *De domesticatie van het noodlot* (Kampen and Antwerp, 2006).
[25] See, for example, Edward Tenner's sobering *Why Things Bite Back*.
[26] This observation is confirmed by Peter Hall: 'A notable fact: during the third and fourth Kondratieffs, when information technology was first developed and then diffused throughout the world, no reduction in travel or face-to-face contact was ever observed. On the contrary, innovations in telecommunications were always paralleled by innovations in transport technology' (Peter Hall, *Cities in Civilisation: Culture, Innovation, and Urban Order*, London, 1998, p. 962).

tier properties I have already defined.

Zapping text The hypertextual nature of the Internet has focused much attention on readers' freedom to make their own way through the information provided. This is, however, not just freedom to discover a personal path through one text, but also to click away from that text to another text or website more quickly. Zapping as a concept may derive from television, it is highly applicable too to the world of digital text. Technologically this is stimulated by two phenomena: the clickable link and the docuverse as one single information space. (In addition, the poor quality of present-day computer screens is also likely to contribute to the tendency to read shorter pieces of text.)

In addition to the readers' awareness of the ever-beckoning vistas beyond their current position in the docuverse there is also the constant competition for attention *in the same digital space* from a plethora of other than textual mediums, including games, video, and music.[27] The printed book is itself a medium in isolation, and it fosters in turn a private communion with the text. Such dedicated form of reading is almost impossible to maintain when digital text occurs in the context of other digital modalities.[28]

In reply to changing reading processes writing changes, too. Notably, shorter text units make their appearance, and existing text is frequently reused, made easier by widespread cut-and-paste habits.

As an example of the kind of 'morbid symptoms' that according to Birkerts typically make their appearance in a 'interregnum', he observes rather mournfully how people discard the familiar forms of book-based literacy. By that he means a literacy based on the linear order of the book, which is 'bound to logic by the imperatives of syntax'.[29] In the 'electronic order', which has yet to prove itself, on the other hand,

> the visual and nonvisual technology in every way encourages in the user a heightened and ever-changing awareness of the present. It works against historical perception, which must depend on the inimical notions of logic and sequential succession. If the print medium exalts the word, fixing it into permanence, the electronic counterpart reduces it to a signal, a means to an end.[30]

The shorter length of the texts that are being read, in conjunction with the greater freedom of readers to find their own path, and the constant distractions offered by other mediums and applications in the same digital space, leads to circumstances that are in any case less

---

[27] This applies perhaps most forcefully to the younger generation. See, for example, Wim Veen, *Flexibel onderwijs voor nieuwe generaties studerenden* (Flexible education for new generations of students), inaugural lecture (Delft, 2000), and W. Veen, F. Jacobs, *Leren van jongeren: Een literatuuronderzoek naar nieuwe geletterdheid* (Young people learning: A literature search for new literacy), Surf-reeks nr. 10, nd [2005], www.surffoundation.nl/download/Leren_van_jongeren.pdf.

[28] The extent of young people's ability to multi-task effectively remains uncertain. In the face of a great deal of optimism (exemplified by Wim Veen among others; see above) neuroscientist Susan Greenfield asserts that 'we have the potential to multitask whilst listening, but not whilst reading' (*Tomorrow's People*, p. 58). Her doubt about the efficiency of multitasking is shared by other neuroscientists, such as Eyal Ophira, Clifford Nass, and Anthony D. Wagner, 'Cognitive control in media multitaskers', *Proceedings of the National Academy of Sciences of the United States of America* 106, 37 (15 September 2009), pp. 15583-87.

[29] Birkerts, *The Gutenberg Elegies*, p. 122.

[30] Birkerts, *The Gutenberg Elegies,* p. 123.

conducive to discursive argumentation.

Drowning in seas of text The first reactions to the Internet gave, understandably, little attention to the massive dimension of the medium. The Internet had been planned as a comprehensive network, but only with the then small group of scientific users in mind. The long-term implications of the non-hierarchical nature of the network for broad access, not only for the consumption but also for the publication of digital text, were not foreseen. The result has become all too prominently visible and goes under the general nomer of information overload. That is the term used to characterise the experience of an overwhelming amount of information, which is not only increasing daily, but in which everything is linked to everything else as well. This information does not have the limitations of the conventional publishing process. The book as a physical product is the result of a process of careful consideration—if only for economic reasons—by the author and publisher of the relative merit of all materials that could in principle be eligible for inclusion. The book as artifact demands from the author a completed manuscript of a certain, always limited size. Moreover, sound or moving images have no place in it, and use of color is limited by financial considerations.[31]

The history of information overload is of course much longer. In the oral tradition redundancy does not exist, in the sense that each utterance is relevant in its context.[32] Also few people will have suffered from information overload in the world of the manuscript: the period was characterised by intensive rather than extensive reading. Moreover, there were not many libraries with large collections and access to them was limited. The production of new manuscripts was simply by demand. It makes sense to assume that the abundance began with the printing press. Not only did the number of books increase rapidly, but printers printed first, and only then looked for a market for their products. A book like *Theatrum humanae vitae* by Theodore Zwinger, professor of medicine at the University of Basel, is a telling manifestation of the problem that confronted early modern scholars dealing with the growing production of the printing press. The work is an anthology of notes on an unlikely variety of topics, for which the author drew from a huge number of books. His endeavour first took the shape of a book of 1400 pages, published in 1565, but the posthumous 1604 edition, edited by his son, already ran to 5000 folio pages of closely set double columns.[33] Its popularity (five editions in forty years, and subsequent editions into the eighteenth century) demonstrates the use for such a work. It presented the notes that someone who wanted to stay *au fait* would have been able to make himself if only he had had time to plough through the steady stream of new publications coming off the press.

The rapid increase in the amount of information as a result of the almost threshold-free publishing opportunities offered by the Internet confronts us even more pressingly

---

[31] It is of course quite conceivable that the zap effect will on average promote the shortening rather than extension of texts. However, in principle, the freedom of variable length increases, and so does, more importantly, the total amount of information, regardless of the length of individual chunks.

[32] As Ong points out, *within* the utterance redundancy does occur, but this is redundancy of a very different kind. Repetition and variation are aimed at enhancing the success of communication (Ong, *Orality and Literacy*, pp. 39-41).

[33] See Walter J. Ong, *Interfaces of the Word: Studies in the Evolution of Consciousness and Culture*, Ithaca and London, 1977, pp. 171-81.

with the same problem. This is not just a matter of bulk. The quality and relevance of this increased flow of information is another point of concern. In *Amusing Ourselves to Death* Neil Postman has observed the implications of the international telegraph traffic in the second half of the nineteenth century for the increase in newspaper news. Much of it was of little relevance to its readers, having no bearing on their daily existence. However, news was regarded primarily as a source of amusement, and people subjected themselves quite voluntarily to this additional supply of pointless information.[34] The apparent lack of discernment of 'the masses' led to a major intellectual distaste for the democratisation of knowledge and information in the early twentieth century.[35] Somewhat similar responses to the increasing flow of not always impeccable information on the web can be heard today.[36]

Incidentally, there are many things Internet users plagued by information overload can do to help them to select the material they really want to know about. The relevance of search results might be improved, for example, by more advanced types of searching than those offered by Google's 'I'm feeling lucky' search field, through the use of alerting services, RSS feeds, tagging, social bookmarks, and so on.

Fragmentation In addition to the sheer quantitative growth of information, there is a growing tendency toward fragmentation of the textual space. On one hand, there is a boom in the number of discrete units of information, on the other hand their availability in the information space of the Internet means that they can be read (and potentially made relevant) in an unprecedented number of lines of argumentation and other contexts. The road up in the hierarchy from data through information to knowledge (not to mention wisdom) is also a process of contextualisation. Short meaningful text fragments on the Web can be reused in a new context, such as a blog. For example, there is a tendency to produce more scholarly 'semi-manufactures'. That is to say that collections of data are published with the intention to serve as a basis for interpretation by others. The infinite number of potential combinations of discrete units of information contributes significantly to the sense of information overload.[37]

The growth of the knowledge space With its previously listed second-tier technological properties the Internet has given new impetus to an old instinct: the gathering of knowledge. The ease and low cost with which new information can be brought online, and the possibility to bring together text virtually from widely dispersed locations has given the human encyclopaedic tendency a new impetus. The Project Gutenberg (1971)

---

[34] Neil Postman, *Amusing Ourselves to Death*, pp. 65-72.
[35] The most famous exponents of European culture pessimism of that time evincing this attitude are undoubtedly Ortega y Gasset and Oswald Spengler. In *The Intellectuals and the Masses: Pride and Prejudice among the Literary Intelligentsia*, 1880-1939 (London, 1992) John Carey gives an entertaining if somewhat caricatural portrait of the conflict between intellectuals and the masses in Great Britain, where it was exacerbated by the well-developed class consciousness of British society.
[36] One is that of Andrew Keen, which I will discuss further below.
[37] Cf. the observation of Jos de Mul that 'The greater the freedom of choice, the greater the uncertainty and hence the entropy … Information overload is not just about the quantity of the information, but also about the fact that the different and frequently conflicting messages increase our uncertainty about the state of affairs' (*Cyberspace Odyssee*, Kampen, 2002, p. 144; my translation).

and its international followers (including the Dutch counterpart, the Project Laurens Janszoon Coster) are semi-scholarly initiatives with the aim of creating and distributing as many 'e-texts' of classical works as possible. In addition, there are various scholarly initiatives, in the form of so-called 'electronic text centres', of which those of the universities of Oxford and Virginia are perhaps the best known. As the publishing possibilities of the Internet increased also outside the scientific world, individual initiatives began to flourish also.

The scholarly and private initiatives to fill the digital libraries dwindle in size compared with recent initiatives that, somewhat unexpectedly, come from commercial quarters. The largest contribution to the universal digital library is now being provided by search giant Google. The profitability of Google's search engine is directly related to the degree in which it is regarded by the user as a better finding machine than the competition. Especially non-scholarly and younger users have higher expectations than most of what they think they are able, or ought to be able, to find in the digital information space. Thus they confirm daily what the director of the Dutch Koninklijke Bibliotheek, the national library of the Netherlands, said in 2005: that non-digital books will soon simply no longerbe visible.[38] So if Google wants to maintain its existing advantage, it is in the commercial interest of the company to bring online not only new but also existing texts. This is precisely what is happening, and at a frantic pace.

In the course of history the question of which texts were worth perpetuating has had to be asked with the advent of every new medium. This happened at the invention of writing, the invention of the codex, and the invention of the printing press. Now the same question is again at issue. But there is one important difference. In a library printed works were stored and used besides manuscripts. They were described in the same manner, to be found in the same library catalogue and to be consulted in the same library. On the Internet, encouraged by Google, people search more and more just on content. That way any manuscripts and printed works that are not digitised disappear from our ken. That makes the digitisation decision much more urgent than previous similar decisions in the transition from book to codex, or codex (and book) to print.[39] With the creation of digital copies in any case it is not necessary to resort to such measures as Ptolemy III is reputed to have taken to expand the collection of the Library of Alexandria: having his customs officers seize any books on board of ships that entered the harbour. That does not mean that there are no problems collecting texts in digital form, but these are mostly in the area of intellectual property rights rather than material ownership. One of the biggest constraints here is that of the copyright protection of books up to seventy years after the death of the author. This threatens to cause a significant lacuna in the universal digital library.

Inspired by Google's good (the size of the scanning operations), but also

---

[38] Juurd Eijsvoogel, 'De stelling van Wim van Drimmelen: een boek dat niet gedigitaliseerd is, bestaat straks niet meer' (The thesis of Wim van Drimmelen: a book that has not been digitised soon won't exist any more', *NRC-Handelsblad,* 10 December 2005.

[39] Leaving such decisions in the hands of commercial companies like Google is not without risk. The Google digitisation project has little if any programmatic foundation. Decisions are mainly led by pragmatic considerations such as the accidental presence of books in contracted libraries, and restrictions set by copyright legislation.

disappointing (the quality of the scans) example, its commercial competitors and library consortia have also entered the mass digitisation fray. All these digital versions of existing books end up in the same ever swelling information space that is now also being filled with film, TV, radio, archives, blogs, forums, online magazines, newspapers and scholarly journals. How large that information space is now, is unknown. The size of the Web can no longer be expressed in numbers of Web pages; so much information is now held increasingly in databases, yielding their contents only in response to a user query. At any rate, despite the enormous energies being lavished on digitisation programmes around the globe, involving information from all disciplines and including our vast cultural heritage, digital information still represents no more than a tiny percentage of the records created through the ages. This percentage will grow at a rapid rate, not only as a result of mega projects such as Google's, but also because the sheer amount of information that is being produced from day to day increases at an exponential rate, and most of it is now created digitally.

However, difficult areas remain. In the analogue world, the books in which our knowledge is stored exist in the same space as the textual archives and special collections of libraries with their correspondences, notes, reports on which much of that knowledge is based. In the docuverse nice successes are being achieved with large-scale digitisation of books and making them machine readable. But digitising manuscript materials remains handiwork. Handwriting recognition remains fraught with difficulties. It is not obvious that it will be possible to make the same digitisation effort in the world of special collections and archives as in that of printed books. For the time being, we live—and perform our scholarly duties—in a dualistic universe where vast dust-ridden archives of analogue paper exist side by side with extensive digital repositories. We will continue to need both for some time to come. But the analogue generation dies off and the new generation knows only what is digital, endangering long-term historical awareness.

Private domain becomes public domain After the invention of the printing press, a distinction grew between the manuscript as at most a semi-public expression and the press as a way of making texts fully public. The long transitional period during which such a distinction became more broadly meaningful is very well documented.[40] In the digital domain the difference between private and public has become virtually meaningless. On the evidence of the exhibitionistic ease with which people share intimate details about their personal lives on TV or via the use of their mobile phone in public places this must be at least in part a broader social development. But it must be to a large extent due to the fact that the personal and the public co-exist without clear boundaries in the same information space. Even if one puts information on the net without the deliberate intention of sharing it, chances are that sooner or later it is found by a search engine. Also, others may consider an intention to limit the publicness of certain materials less important than the author, and whether or not with evil intentions, be more readily disposed to relieve it from its obscurity. Video clips and photos are continuously posted on YouTube and Flickr against the wishes of the maker and / or the person portrayed. As many have found to their

[40] For example by Harold Love (*The Culture and Commerce of Texts*) and by Gerd Dicke and Klaus Grubmüller (eds., *Die Gleichzeitigkeit von Handschrift und Buchdruck*).

chagrin, it is virtually impossible to undo such acts of making public.

But mostly privacy and obscurity are precisely not the intent. To be findable for others is the goal from the outset—even when it comes to material about which others might perhaps be inclined to raise their eyebrows. To be able to be found and to be connected to other texts published in the huge information space that is the Internet is a major incentive to place texts and other materials there, regardless of ownership problems. Just as to be called on the mobile phone is a sign of popularity, or at least proof of existence, so is to be found on the internet.

The social web Not only the medial content itself but also all meta-information about it is located in the same information space where the Universal Machine's massive and ever expanding variety of applications can be applied to it. This fact, joined with the two-way architecture of Internet traffic has created the conditions for the so-called Web 2.0 environment. Web 2.0 is, despite the somewhat vague meaning, a concise way to refer to a number of related phenomena that have occurred over the past few years—since about 2003. These are phenomena like the development of Web communities and services, such as social-networking sites,[41] blogs,[42] and so-called folksonomies,[43] and wikis, about which more below. The perception of the Web is thus changed from a structure in which a small number of providers publish Web sites for a large number of users, to a platform where all users can publish their own knowledge and opinions. The focus is now on the opportunities for participation and interactivity.

It is often stressed that the innovation that the term Web 2.0 suggests lies in a changing use of existing technologies more than in the development of fundamentally new technologies.[44] Anyway, in conjunction with the huge growth in the number of users, a new way of thinking about the web has come about.[45] And it is not a matter of just thinking: the interest in social mediums is growing rapidly.[46]

One of the best-known mechanisms to improve the quality of the collective knowledge on offer is without doubt the wiki. The principle of the wiki was invented not

---

[41] On the so-called social-networking sites, such as Facebook, MySpace, and LinkedIn, users can create personal profiles, with biographical data, personal preferences and the relationships that maintained, virtually or in real life, with others. Thus, virtual communities can come into being, on the basis of shared personal preferences, work and so on.

[42] Blogs (short for web logs) are a popular publication platform for personal views on any conceivable subject. The shape resembles that of the diary, in which notes are organized primarily by date. By granting descriptive tags the notes can also be grouped thematically. From personal confidences to reflections on the news or the results of scientific research everything can be found in blog form. Most blogs are maintained by one person, but there are also collective blogs.

[43] Folksonomies are taxonomies that result from non-experts collaboratively assigning descriptive tags to digital materials. On http://del.icio.us, for example, users can publish their bookmark collection and assign tags to Web sites. Folksonomies are often visualized in so-called 'tag clouds', which show relations between tags.

[44] For example, Tim Berners-Lee in 'developerWorks Interviews: Tim Berners-Lee', 22 August 2006, http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html.

[45] However radical the social and economic impact of Web 2.0 can be is demonstrated by C. Pascu *et al.* in 'The Potential Disruptive Impact of Internet 2 Based Technologies', *First Monday* 12, 3 (5 March 2007), http://www.firstmonday.org/issues/issue12_3/pascu/index.html.

[46] Market research by Hitwise in 2006-2007 shows an increase of 668% (http://www.businessweek.com/the_thread/blogspotting/archives/pdf/Tancer%20Web2expo1.pdf). However, active participation remains still far behind the passive use of the web.

only to mobilise the 'wisdom of the crowd' but also to regulate it. The way it works is that everyone can both make a contribution themselves and correct the contribution of someone else. Wikis do not necessarily have to be publicly accessible. Closed wikis provide an excellent infrastructure for interest groups consisting of experts and/or non-experts to pool and organise their collective knowledge. This self-regulatory environment, as in the case of open source software, is designed to lead to better quality. Eventually, in the process of writing and rewriting, the best information should float to the top. Whether or not that is the case in the publicly accessible wikis, is not always clear.[47] An objective measure of the quality of software is rather easy to establish: for example, whether or not the software functions as advertised, and the reliability with which it does so; its robustness, expressed in the duration of the period that it continues to work without failure; security, expressed in the degree to which it is resistant to hacking. But in the atmosphere of encyclopaedic contributions to, for instance, Wikipedia less objective standards apply. Especially political and aesthetic motives appear to play a strong role in the wording of the 'facts'. In the continuing tug-of-war 'corrected' versions of articles keep being 'improved'. In a more general sense, this is the problem of the Internet already identified: the qualified expert and the self-appointed expert move in the same information space. It is up to the user to discern the difference and that is no easy task.

The contribution of non-experts, also called 'the wisdom of the crowds', is particularly interesting in cases where automation is not possible. In the case of the above-mentioned applications the aim is for the user to make a conscious contribution. In the case of the various projects that Luis von Ahn has devised under the name of 'human computation', the Internet-user is not even aware what contribution he makes (or even that he is making one).[48] Since they are playing games that are entertaining enough in their own right the users need no further motivation. Von Ahn's games are a good example of collective power when it comes to contributing to digital knowledge creation.

With the increasing involvement of the non-expert, however, again a part of the interpretive burden is shifting from the instigator of the communication to its recipient. As noted in Chapter 3, that was already one of Plato's objections against the written compared to the spoken word.[49] More generally, the emphasis in digital communications is shifting increasingly from the transfer of knowledge (where readers can in principle remain passive, trusting their source) to the transfer of information which can lead to knowledge. The reader gets a heavier responsibility for at least the validation of knowledge, but more often also for its constitution. That shift is exacerbated by several other developments, such as the zap-like nature of the medium and the changing context in which text can be

---

[47] In 2005 the British journal *Nature* examined the reliability of a number of scientific entries from both Wikipedia and the *Encyclopedia Britannica*. The two publications were a close match. 'Only eight serious errors, such as misinterpretations of important concepts, were detected in the pairs of articles reviewed, four from each encyclopedia. But reviewers also found many factual errors, omissions or misleading statements: 162 and 123 in Wikipedia and Britannica, respectively' (Jim Giles, 'Internet encyclopaedias go head to head', *Nature* 438, 7070 (15 December 2005), pp. 900-901).

[48] These are examples of collective tagging; see http://www.cs.cmu.edu/~biglou.

[49] On the other hand, in the digital world this shift of the interpretation to a later point in the Communications Circuit is somewhat offset by the previously reported shift to an earlier moment when findings are shared: by people in general, for example through blogs in the public space; by scholars in their own scholarly circuit.

deployed. The argumentational and more generally discursive nature of the linear analogue text seems to be taking an ever less prominent place in the digital environment.

Text of typing monkeys and other quality issues The social developments sketched so far pose especially urgent questions about the quality of the digital offerings. That average quality is often perceived as low has various causes. Firstly, there is the already mentioned low threshold to active participation in the medium, which has led to an avalanche of publications of the most diverse kinds: from scholarly data collections to personal effusions of an often toe-curling nature. Poor average quality is the other side of the coin of ease of access and lack of control. Never before in history has the ability of humans to express themselves in a public medium seen such explosive growth. Without any economic disincentive, social inhibitions or political control, anyone can publish what they want. The traditional distinction between user and publisher is vanishing: the medium is truly in the hands of the user.[50]

Recently, Andrew Keen, a reformed internet entrepreneur, has severely criticised what he regards as the appalling quality of the digital information space. In his book *The Cult of the Amateur: How Today's Internet Is Killing Our Culture*, he compares the legions of self-publishing amateur writers, filmmakers, and musicians with typing monkeys, who together produce an endless and depressing amount of mediocrity. Keen's reaction to the way the Internet has developed in recent times reflects the widely shared view that the democratic idea on which the Internet was founded has tragically derailed.

A related issue is that of the perceived lack of moral tone. It is true that the level of rantings and ravings is high. The ease with which people tend to give their uncensored opinions seems to invite eye-for-an-eye, tooth-for-a-tooth reactions. While speculation about the cause abounds, it must be admitted that the medium's anonymity is a very likely contributing factor.

A very different quality issue concerns the precision, or rather the lack of it, with which existing information is being digitised. The new—born digital—information is not always reliable, but also the transfer of existing information is not always good. The reliability of Google's scanning and OCR, for example, is often criticised. The absence of the traditional economic model for for the publication of this information is likely to be one cause.

This, and such a phenomenon as text disappearing, add to the perception that the digital environment does not live up to the standards of the book that is so familiar that it has come to serve as our gold standard. Its instability is one of the most striking features of digital text. Digital texts are often here one day and gone the next. Even more easily than a piece of paper—let alone an entire book—a digital text can be destroyed for ever. Even if a copy has been preserved somewhere, the integrity of that copy is not guaranteed. Just as easily—but much more difficult for the reader to see—the text on a website may have been changed without any account of the changes being rendered. (In many Web pages published under some form of editorial responsibility changes are well documented, as in

[50] One of the interesting side affects of the unbridled popular access to the digital writing space is that it may well have increased writing activity overall. Blogging, MSN messaging, pdating FaceBook pages, sending SMSs are all social forms of writing that hardly had an analogue equivalent.

the case of Wikipedia. But publishing on the Internet is not reserved for responsible publishers.) The ease with which a digital text can be changed unfortunately goes hand in hand with the invisibility of the change. Thus readers are hard put to know which version of a text they are reading.

Coming from a typographical world, we are too easily tempted to experience instability mainly as a form of unreliability or untrustworthiness. It may be necessary to take a slightly different view of the phenomenon of instability. It is after all precisely by the grace of that property of instability that the entire digital docuverse exists at all. Moreover, the network preserves much, too, which inadvertently makes for greater permanence than is often assumed. Saving occurs both intentionally and unintentionally. In the first category belong, for example, Internet archiving projects and the creation (for whatever reason) of local copies of material published elsewhere on the Web; the second concerns the copies generated on the client computer or in caches on proxy servers as a result of consultation and distribution. As was already argued in 'The end of the copy' (one of the second-tier properties of digital text) these copies are not functionally distinguishable from the original. However, the fact that those files are no longer in their canonical location (URL) does cast doubt on their integrity. The least we can do is learn to live with this instability, which is not only useful to a degree that we don't always recognise, it is after all also technologically determined. That is not always easy to accept for *homo typographicus*. It is precisely our ever-growing historical awareness (which is linked with our medium use) that ensures that we are fully aware of the importance of our textual record to our identity.

To discard documents deliberately is one thing; letting them languish (as a result of 'link rot' or not migrating outdated formats) is something else. Yet this too is a typical problem of digital technology. Unless actively resisted, digital ageing is inevitable. Digital ageing has two main causes: the physical deterioration of the physical supports, and digital obsolescence, from changes in hardware, storage format and operating system and application software. There are strategies to deal with it, such as renewal of the physical substrate and / or making copies; the migration of file formats or operating systems; and the emulation of hardware platforms, operating systems or applications. Conscious preservation is without doubt one of the biggest challenges of the digital era. The efforts made to achieve digital longevity—preservation of digital documents and ensuring access to them in the long term—in the constantly evolving hardware and software environment may be great, but the speed of change—and thereby the process of ageing—is increasing all the time.

Status of digital information All these aspects of quality and ephemerality jointly lead to profound questions regarding the value and status of digital information. The deterioration of the status of text in a digital environment is an insidious but not necessarily new phenomenon. With every cheap edition of the classics ever published something of the 'aura' of the original artwork was lost. Indeed, the whole idea of 'the original' has been severely compromised by the invention of printing. Yet the digital medium has marginalised the notion of the original even further. In a sense, this is paradoxical. After all, digital copies can not be distinguished from 'the original'. But even if

the same quality can be ascribed to each digital copy as to the 'original', the value of both is nevertheless irrevocably lower, on the same economic principle that makes a mass paperback less valuable than a copy of a book composed and printed by hand in a limited edition. For this devaluation to occur no actual copies even have to be made. The awareness that the 'copying press' built into the document has the potential to keep running indefinitely suffices. This, combined with the already mentioned instability and low average quality of what is on offer on the Internet, results in an inevitable sense of expendability. In other words, the content is tainted by its context.[51]

Internationalisation The Internet is global and in principle knows no boundaries. The market for information, commercial or non-commercial, is by definition largely global. One of the ways in which this becomes evident is through the use of English as a *lingua franca*. Not only is English the language of the communication itself, it is also the technical meta-language that makes communication possible. HTML and XHTML—the code languages that make the Web possible—can only use English tag names because browsers only read English. In principle it is possible to use the meta language XML to write markup languages in any human language that can be displayed in Unicode. For practical reasons, however, XML is predominantly written and used in English, precisely because XML is designed as an interchange format. The use of, for example, Arab element names would mean that not only could these not be read by people who don't speak Arabic, but also not by non-Arab software and more generally not by computers that are not equipped with an Arabic writing system. A certain cultural homogenisation is the inevitable result.

Also in the case of non-linguistic modalities, which play a growing role on the Web, there is a certain homogenisation in evidence. Communication is often supported with sound, colour and graphics which are less language and culture-bound than text. The web interface too mutes cultural differences. While a Japanese book differs from a Dutch book by needing to be read from back to front, a Japanese website does not have to be read from bottom to top. Because the technological constraints (such as the use of HTML) are equal for any language, the entire typographic and multimedia information space tends towards international convergence.

Globalisation leads to new social relations, but also to other economic (commercial) relations (see further 'Communities' below). The instruments for international regulation lag seriously behind this tendency. On the one hand local decisions may have international implications; on the other hand local laws are often inadequate when it comes to a medium that is by definition international.

Communities Someone who reads isolates him- or herself and simultaneously is part of a community of readers, in particular readers of the same book. That readers are alone

[51] It is possible that while the significance of the book in the transfer of information (as opposed to leisure) decreases, paradoxically the iconic significance of the book as a cultural medium increases. This would in any case be one possible explanation for the fact that the number of books sold is increasing (for the Netherlands see e.g., the GfK *Jaargids 2008*, Chapter 19, 'Lezen en gamen favoriete vrijetijdsbesteding' (Reading and gaming favourite leisure pursuits), while the time spent actually reading decreases (see e.g., Frank Huysmans, 'De openbare bibliotheek in Nederland en de veranderende leescultuur sinds 1975' (The public library in the Netherlands and the changing reading culture since 1975), in *Jaarboek voor Nederlandse Boekgeschiedenis* 14 (2007), pp. 179-92).

with the text they read is often observed to be a characteristic that distinguishes written communication from oral communication.[52] This is not only a social consequence of the technology, it is also the result of the demand that the physical activity of reading makes on the brain.[53] But in addition to the isolation of the direct physical environment that is created by reading, the book also offers access to a virtual spiritual community. As long as reading was the privilege of a small group in society this sense of community did not even have to be about a particular text. Merely being able to read, meant being part of a small, select social group. Now that reading is no longer reserved for an elite, the actual content—a specific text or genre—plays a larger role. Such a community does not have to exist synchronously; the reading experience can be shared with people at another place and time. Like a book, a computer screen isolates the reader of a digital text from his or her immediate surroundings. But just as with a book, through a computer screen the reader can also participate in a virtual community. In fact the screen—provided that it is part of a computer that is connected to the Internet—offers even more opportunities for that. The screen provides a vista of a communicative world without borders. The sense of community lies not only in the experience of a virtual connectedness, but also in actual and interactive communication. This may also find other than written expression, for example, through moving images (webcam) or audio (Voice over IP or VoIP). However, the screen can in some respects limit the sense of community, too. It is a technological property of printing that it is only profitable in a certain printrun. The idea of wider dissemination is a core property of the technology of printing. It is precisely one of the salient characteristics of the digital medium that there is no minimum 'edition size'. That is to say that the low mean size of the digital audience may be much lower than that of the print audience. A tension thus becomes discernible between the (potential) global reach of the Internet as a medium on the one hand, and on the other hand the medium's possibility to reach small, previously not economically or technologically accessible communities: a niche audience. Digital publication can respond to a very individualised interest much better than print can. 'Publishing' has always been regarded as being a form of broadcasting, but it is starting to make the transition towards narrowcasting even now. Online database publishing of scholarly periodicals, for example, allows such narrowcasting much more readily than print-based publication of the same material. The limits to what individual readers could find are not in fact determined by lack of material to their liking, but by the scarcity of their time and attention: the so-called attention economy.

The narrowcasting capabilities of the Internet will obviously have an impact on the sense of community. If it is assumed that communality is an important basis for certain processes (such as, notably, democratic ones) that have a utility or are prerequisite for the way society functions, this could have all sorts of social consequences. However this may be, the global nature of the Internet puts pressure on the notion that 'imagined communities' are supposed to be based on geographical closeness. If the public sphere of shared political, economic and cultural interests will be replaced by imagined communities on the basis of purely personal interest, this may benefit the individual, but not necessarily the community of which they are part geographically.

---

[52] See, for example, Ong, *Orality and Literacy*, p. 74.
[53] Cf. Greenfield, *Tomorrow's People,* p. 58.

The interface 'Interface' is typically defined as human–machine interaction. In particular, the term is mostly used for the way humans can control the computer's software: the operating system and applications. All modern computer interfaces now being of the Graphic User Interface (GUI) type, they use Windows, Icons, Mouse and Pulldown menus (WIMP). This digital interface is almost always quite emphatically present. The screen on which the digital text and other information is presented to the user always provides a significant number of buttons, icons and pull-down menus. Depending on the computer's operating system and the application being used, this arsenal of controls on two different levels (of operating system and applications) can, to a greater or lesser extent, form a unit.

But the operation of the software is just one part of the human–computer interaction. Two further elements which contribute to ensuring access to digital text need consideration. One element is the representation of the text itself (within the operating and application software) and the other is the decisive contribution made by the hardware.

As regards the representation of the text itself presentation software (mainly browsers) offers quite modest possibilities for variation. As far as the World Wide Web is concerned, the bandwidth of this variation is fully determined by what is preprogrammed in browsers. There (X)HTML and XML are the standard exchange languages. It is precisely the requirements of interchange that impose restrictions on the presentation, in spite of all stylesheets possibilities. At first glance this seems to apply to a lesser degree to another exchange format, the popular portable document format (PDF), which perfectly emulates the typographical capabilities of print. PDF was devised to get better control over the instability of digital form and it has proved a very effective solution. PDF files are platform-independent and describe the typographical shape of every page with exacting precision. Yet certain restrictions persist, perhaps precisely because of the implicit reference the format makes to the typographical capabilities of printing. The user will always remain very conscious of the screen as a kind of frame through which he looks at an image of typographically shaped text. This effect is further strengthened by the fact that the operating system usually remains—more or less emphatically—present in the presentation of the text.

The screen as a frame draws attention to the determining role that the hardware plays in the interface. In practice, a user usually works with a single screen, on which all digital information is displayed. Therefore, it is always the properties of this single screen that determine the representation of all information, regardless of the potentially very diverse nature of that information. This screen is always rectangular, almost always landscape-oriented, full colour, and backlit. Moreover current display technology is characterised by a low resolution, which also severely limits the typographical possibilities. Few fonts prove properly legible on the screen. Instead of the unity between content and form that occurs in the case of printed information (which is determined by the publisher) the presentation of digital text will thus be adapted to and determined by the hardware. All this makes for a high degree of uniformity. Besides the computer screen there are screens of mobile phones, PDAs, e-book readers and so on. But here again, the shape of the screen remains the same per hardware category whatever the information that is being displayed,

and so largely determines the representation. Only the newest generation of e-book readers uses a fundamentally different technology. Their black-and-white screen is reflective and has a high resolution, but this technology is only in its infancy. Here at least the role of the operating system and application software are potentially less intrusive, since these are basically single-purpose devices (which is, however, at the same time their besetting weakness).

The author or publisher of a digital text therefore has relatively little control over the form in which it reaches the reader. On the one hand this is the result of the restrictions imposed by the hardware and software, leaving only limited influence on the way in which it is displayed on the screen of an individual user. This low bandwidth makes for a certain homogeneity of form. All digital texts, regardless of provenance or quality, look identical. On the other hand individual readers—within the same limits—have the freedom to determine what form they want to impose on the enormous diversity of as yet 'unformed' content to which they have access. Users can adjust the interface to their own tastes and preferences through both pre-programmed 'skins' and fully self-programmable elements. The limited control the publisher of digital text has over its visual appearance, added to the consumer's own influence, makes for a situation that is diametrically opposed to that of the book as a medium. The book features an unbreakable nexus between content and form, the form being fully determined by the publisher. From the moment that form is fixed, the text can be relied on to present always the same content in the same form, offering exactly the same presentation for every reader everywhere. It is this characteristic of printing which has enabled all sorts of subtle typographical distinctions to become so firmly lodged in the collective (un)consciousness of *homo typographicus*.

It is instructive to carry the comparison with the physical book a little further. It works in two directions. Armed with our new knowledge of computers it appears that the book—without our knowing it—also already had an 'interface'. Just as in human–machine interaction the interface is an informational layer that offers the user functional access to the 'content' of the computer, the book as a reading machine also has functional properties that offer the reader access to its content. Together these form the 'user interface' of the book. This includes, materially, the form of the codex (which unlike the scroll allows browsing) with its characteristic rectangles of text, surrounded by white; the reading direction, both of the lines on the pages and of the pages in the book; the presence of such ordering elements as page numbers, table of contents and index; the canonical order of the elements that make up a book; and the presence of identifying title, author and other publication data on the binding, the cover, the title page, the colophon or reverse title page. In a less material sense it also includes the use of these elements in reference systems such as footnotes and bibliographies. Also characteristic of printed texts is of course that their physical form as a unit coincides with the unity of content, while digital texts are submerged in the docuverse.

The digital text forms that are becoming so important have freshly opened our eyes to the by now so intuitive way in which books provide access to the text. The interface of the book is so self-evident that instructions for its use are unnecessary. It is its predictability and familiarity that makes the book's canonical user interface so intuitive. The history of page numbering in Chapter 3 demonstrated that this predictability and

familiarity arose only slowly. Gradually the concept of the book has become increasingly predictable and reliable. It goes without saying that you always know where you are in the text, by the presence of page numbers, headers or footers, but also by the thickness of the part of the book already read in relation to the unread part. Between beginning and end of the book the page numbers go up. The thickness of a book allows you to estimate—roughly—the extent of its content—something that is made possible by the conventions that have come to determine the choice of format of the book, the type size and the proportional relationships between them. Our familiarity with the conventions is such that a single look at the physical form allows us to determine very reliably the nature of any printed text. Drama or prose, children's book or scholarship, entertainment or newspaper: they can be easily recognised from a distance without any conscious effort or thought.

Conversely, looing at the digital textual medium armed with a thorough knowledge of the book, the clumsy nature of the digital interface contrasts sharply with that of the reading machine that the book has become in the course of time. We are so familiar with the interface of the book that we are hardly aware any more of the intuitive way it works. In the case of the digital textual medium, it is often still necessary to take recourse to such instructions as 'Click here to go to the next page'.

The book interface offers achievements which have come to be appreciated as a great good. As a result of both the extent to which we are conditioned by our familiarity with the book and the continuity in function between book and digital text, we expect involuntarily and unconsciously to find in the digital medium some of the properties of the book interface. What are then found lacking in the comparison are things like a digital alternative to the concept of the page, and a typographical form language that is subtle enough to be able to recognise genres. But also, for example, the fact that a text once printed no longer changes, appears to be a useful property that is often sorely missed. Not only does the analogue text as a whole always remain equal to itself, but a certain passage in a book always remains firmly anchored to a physical location in its layout. For example, a reader may even remember to have come across a certain remarkable observation 'somewhere at the top of a left-hand page'. It is also these properties that enable the impressive granularity of our bibliographic referencing system. This is the type of certainty of the printed text and the comfort they provide that are sought in vain in the digital world.

The whole digital interface bears the heavy stamp of the software and hardware that define it. They restrict the possibilities for familiar interface elements of the book to be translated into the digital environment. But how bad is this? It makes no sense after all to wish ourselves a horseless carriage. It may be a technologically fascinating challenge to create a faithful translation from one interface to the other, but how useful is it to try to transplant every feature of the book interface to the digital environment? Does our future really only lie in the past? Had not we better ensure that the digital interface does justice to the inherent properties of digital textuality? Should the digital interface not exploit the fluidity of the digital text, its non-linearity, the possibilities of non-verbal communication that result from the use of other modalities, instead of treating them as problems to be resolved? For example: what kind of 'meanings' may hyperlinks have and how can they be expressed? How can we enrich the specific digital 'form language', typographically or otherwise? How can users gain a better representation of the nature and extent of the text

that is being access through their screen at a given moment? It is precisely the fluid, virtual form of all digital text that forces the question how its visualisation can best do justice to the content.

From this viewpoint, that the digital interface must do justice to the inherent properties of digitality, it is easy to believe that it is only a matter of time before the interface to digital information begins to crystallise and appear as familiar as the predictably unequivocal interface of the printed book. That would, however, be a dangerous assumption. That notion misunderstands the dynamics of the digital medium. Not only is the digital interface still evolving, with new navigation methods as well as new *standards* for navigation continuing to be developed for the foreseeable future. It is also unlikely that a digital interface will ever be 'finished' in the way the book's interface is. The computer is after all a Universal Machine. In the meantime, the expectations based on our familiarity with the book are not likely to disappear quickly either. For the book itself will not disappear quickly from a society all of whose institutions are intertwined with books and print.

Conclusion

In Chapter 1, I introduced the concept of primary technological properties of mediums and suggested that these may provide an explanation for certain second-tier symptoms, which in turn cause social effects. In Chapter 3, I showed how this mechanism works in the case of the introduction of the printing press. In this chapter I have applied the same approach to the digital textual medium. It is clear that in both cases there are indeed far-reaching social effects that can be traced back to certain primary technological properties. What picture emerges if we look at the social impact of the introduction of the digital textual medium?

Many of the phenomena discussed in this chapter can be placed under the general nomer of the democratisation of the means of publication. For someone who would have got stuck in the era of the printing press, and had missed the development of the digital medium, the current situation would have been utterly inconceivable. As an alternative to the typewriter, authors now have available a computer and a word-processing application. For the publication of their writings they now have that same computer on which they have written them, which serves as an alternative to the triad of printer–publisher–bookseller. The computer-in-a-network—again the same device that runs the word processing application—does in a matter of seconds the hard work of distribution that used to take months. And it does not stop there, because the reactions from readers can be expected to start flowing in—again on the same apparatus—almost immediately. Authors can read these comments themselves, but if they choose, they can offer to share both their work and the responses it evokes with all and sundry—and react again to the responses. That is a form of direct two-way traffic that is totally alien to the world of print.

In one fell swoop the possibility of self-publishing has removed the entire slowly grown system of authorisation that belongs to the traditional publishing process. That includes the control of quality that the publisher used to exercise. At the other 'end' of the communications circuit the same applies. The value of the intermediary role of libraries as

instruments of regulation and ordering and bookshops as instruments of selection is absent in the docuverse. As a result the reader is confronted with materials of very variable quality and diversity, mixing information intended for a larger public with very private information in one textual space. Moreover, all that information is connected: readers need to pick their own route through the abundance of—fragmented—information. For this they can click on hyperlinks, but also search by content with the help of search engines. Either way users can zap their way from one piece of information to the next. The multimedia quality of the Internet encourages that: it is the same consumption behaviour to which the TV viewer is accustomed.

The status of digital text is adversely affected by its varying quality. This is reinforced by the here-today-gone-tomorrow (or at least changed-tomorrow) character of web texts. This means that the digital medium represents a momentous break in the historical evolution towards ever greater permanence of the repository of human knowledge. The flexibility that characterises oral knowledge disappeared when knowledge was put in writing. Instead a degree of permanence and objectivity emerged. In the tradition of manuscript transmission this still remained somewhat limited due to the fact that in the process of copying small adjustments could continue to be made, consciously or unconsciously. With the production of printed matter began an irreversible and unstoppable process of stabilisation and canonisation of knowledge that gradually led to the monumentality that we have come to associate with printing. This process is reversed by the digital medium, placing a heavy interpretive burden on the reader's shoulders and making high demands on consumers' critical faculties.

In addition, the digital medium is not limited to the simple transfer of text. It is hardly even possible to separate the medium and its conventional functions of production, distribution and consumption represented by the communications circuit from what the Universal Machine can do.

The list of social consequences of the inherent technological properties of the digital medium above ended with the instability of the digital interface and the difficulty of translating familiar elements from the interface of the book to the digital environment. This very method, of drawing parallels with other textual mediums, has of course been suggestive of continuity. Indeed in Chapter 1 I myself introduced the concept of continuity in textual transmission, from handwriting to print to digital. I did that primarily on the basis of the *function* of transmission: in all cases the purpose is to transmit texts from one person to another that have in some way been 'inscribed'. In this chapter I have again suggested continuity in many ways, for example through the use of the same term 'user interface' for both books and the digital medium in the last item on the list of social consequences. At the same time, however, every item on the list has shown major discontinuities between the social consequences of print and digital textual transmission. Based on this survey, it must be concluded that the degree of social continuity is much less than one may involuntarily have come to expect from the functional continuity between these two technologies and the fact that all sorts of conceptual correspondences can recognised between them. The conclusion seems inescapable that not only the technological characteristics of books and the digital medium, but also their social consequences are indeed very different.