The Libraryness of Calculative Devices:
Artificially Intelligent Librarians and Their Impact on Information Consumption

Martijn van Otterlo

(int this version may still be modified slightly. Please cite as:


Introduction

Physical books are colourful objects: “[r]egardless of how books are grouped, they do furnish a room” (Petroski, 1999:252). However, libraries are not formed randomly, but it is the owner’s intention to accumulate knowledge on particular topics that forms them. Libraries are deliberate acts: each book contributes something, and all books together form a unique knowledge base, selected, arranged and ordered by a librarian, who has, in essence, only two main tasks.

The first task is the selection of books. The librarian brings together a set of books, through buying, gifting, trading and so on. The librarian’s interests, among those of a wider community of scholars or readers, determine why books get selected. The librarian’s second task is to order the books according to some criteria, and physically arrange them according to that order. Several formal systems have been developed in history to do so. For most personal libraries, ordering is a constant activity due to limited physical space, an expanding collection of books, and changing interests. Rearranging books (e.g., by subject) is a fun but tedious thing to do, and, in the end, the order is never perfectly right (e.g., does Alice in Wonderland go with science or with novels?), and often physical space is lacking
in the wrong places (e.g., subjects spanning exactly one full shelf plus three books: where to put these then?). In general, the order of the library depends on the physical spaces in which it resides. This includes the physical shelves (see Petroski, 1999), but also the library’s building.

**What a famous book hunt teaches us about calculative devices**

I frame information consumption activities, such as reading or searching on the web, as traversing a (digital) library. By way of illustration, consider the famous story ‘The Name of the Rose’ by Umberto Eco (1980), in which, beneath the plot of a murder mystery, a grand story about knowledge unfolds in the context of a cloister library. The main characters find in the blind librarian Jorge von Burgos a notable adversary, who tries to keep them from finding a particular book. They find themselves in an unknown library, with hidden passages, hidden books, and with an order unknown to them. The particular arrangement of books was beyond their capacity to navigate the space and locate the specific text. All of these obstacles were introduced by the ‘evil’ librarian with a purpose: to control what others read and especially what they should not read. The library is a source of knowledge, but it was the librarian who decided the terms of its use.

In addition to physical libraries, which still exist of course, many novel, digital libraries have arisen in which we can search for knowledge. For example, Google Books, containing millions of books, can be seen as a digital library, where books can be ordered on the fly according to many possible criteria. Whereas Google Books is somewhat close to a physical library, we can also talk about general web search as library traversal: one starts at some point and, based on the order of information sources on the web (governed by the search engine), we are given an ordered set of items we can visit (the search results). Just like in Eco’s story, it is essentially the librarian who gets to determine how our traversal through his library comes about. But, unlike in the novel, digital libraries have digital librarians; pieces
of software, or, *algorithms*. Algorithms govern the library via automated selecting and ordering. Digital librarians are *calculative devices* that learn from the information-seeking behaviour of many individuals to instantly *rearrange* the library for a specific individual and her current information needs. Modern search engines *personalise* search results based on previous behaviour, general knowledge and statistical information about many other users. Understanding that particular arrangement of material, and how this reordering happens, is vital for appreciating how much power algorithms have obtained in our society.

To understand how algorithms function as keepers of digital information, one may study algorithms as things with capacities themselves. An alternative developed in this Chapter is to exploit our pre-existing knowledge of physical libraries through the use of a *metaphor* that links it to new contexts. Metaphors are *thought tools* and vital in framing difficult problems to understand them in new ways. Well-known examples from the privacy literature are Orwell’s ‘*Big Brother*’ and Bentham’s ‘*Panopticon*’. Here, I introduce a ‘library metaphor’ to characterise modern information consumption, which I conceive as a traversal of a digital library governed by a digital librarian.

The Chapter proceeds as follows. First, I briefly review some aspects of the overarching topic of privacy as control over information, in which our metaphor finds its place. I then outline fundamental aspects of physical, digital and universal libraries, before moving on to discuss the role and power of digital librarians, mapping out some insights on how to better study them. Overall the Chapter proposes a new means of characterising and understanding the work of algorithms in our world.

**Privacy and control over information**

The ideas I develop here contribute to a wide-ranging contemporary debate on algorithms, surveillance, privacy and control in digital worlds (e.g., Vaidhyanathan,
It is possible to distinguish three types of changes in control over information: access, prediction and manipulation.

The first type of control centres on access. Currently much of the privacy debate is about what Google, Facebook, the NSA and others know about us. Framed as Big Data, information is gathered everywhere, ranging from medical records, to public transport chip cards, and to smart energy meters. Long before the infamous “don’t be evil” slogan by Google, Warner and Stone (1970:146) warned us to “not be naive about it”: “Anyone who has entered into a hire-purchase transaction … should nowadays expect both the personal data he supplied in his application, and the information about his reliability in making the repayments, to be widely available”. Roughly forty years later, information is available in digital form, and privacy violations and abuse of data are common. Data is a commodity and is traded on a large scale. Control over information here is essentially about control of access: who-can-see-which-information.

The second type of control arises when intelligent algorithms are employed, such as with machine learning (Domingos, 2012), or more generally, artificial intelligence (AI) (Nilsson, 2010).

Such algorithms couple modern statistical methods with powerful knowledge representation languages to generate rich prediction models from data and allow for the generation of new or inferred knowledge. Such models are based on information about many individuals, and they can be used to predict traits for
specific individuals. Going beyond the data alone, the secondary use of data was aptly termed by Amoore (2011) ‘data derivatives’. For example, models could (probabilistically) predict whether I would buy a particular book, based on previous purchases and similarity measures between books that are ‘alike’. Elsewhere (van Otterlo, 2013), I have discussed techniques for the generation of models from data. These may contain a generalised rule ‘if a person is tall, it is more likely that it is a male’, which represents a typical pattern in the data. Such a rule may not predict well for every individual, but it does predict well on average. In addition, the rule may be used to predict (a possibly unobserved feature) ‘male’ from (an observed feature) ‘tall’. Models can be utilised to infer more about a specific individual and to statistically predict the behavior of individuals. For example, Schwartz et al. (2013) predict features such as gender, age and location of people based on their language use, whereas Kosinski et al. (2013) predict such traits from Facebook ‘likes’. Many forms of bias are present when generating the models from data, which determine how accurate or confident predictions are.

The third type of control over information deals with the use of prediction models for a purpose, such as surveillance or commercial profit.

Figure 2. Feedback loops and experimentation.
I have outlined (van Otterlo, 2014) how modern computational techniques give rise to communities governed by the principles of *behavioural engineering* as described in ‘*Walden Two*’ by the psychologist B.F. Skinner. The idea is simple, yet very powerful: given a prediction model of individual behaviour a company or government can exploit that model to manipulate the behaviour of large groups of individuals. As claimed by Warner and Stone (1970:124):

> Give the administrator in government or business the use of an integrated national population file ... and you provide him with a powerful tool for interference in private lives, to manipulate, to sell more, to condition, to coerce.

Currently, based on the countless ways people are *measured* in their daily activities, algorithms can automatically build prediction models to do just that. Since this happens in a statistical manner, the manipulation of individuals may be incorrect or inaccurate but, on average, at the level of populations, manipulations can be understood as ‘error tolerant’ and successful. Good examples of the exploitation of models are supermarket loyalty cards and modern political campaigns, where social media is used to target specific sub-populations. Such situations also highlight a possibility for full *feedback loops* (van Otterlo, 2009), in which data collection, model generation and model exploitation are executed in sequence and indefinitely, enabling algorithms to experiment with different settings to see which (kinds of) manipulations work best. For example, many websites present different versions to different users to experiment with layout and information presentation, in order to maximise profits.

The three types of changes I have discussed are ordered by how strong algorithms influence the information environment of a particular individual. Where, in the first case, information is merely gathered, the second deals with using that information for prediction and due to that treating an individual in new ways. One step further, the third setting concerns the direct, intentional manipulation of the information environment of an individual and provides the general setting for what follows.
Libraries as ordered spaces of information

In this section, I discuss two types of libraries, physical and universal, which form two distinct ways to view and organise a collection of data or information sources.

![Image](image_url)

**Figure 3. A librarian in: (left) an ordered, physical library; and (right) – in an unordered, digital or universal library.**

**Physical libraries**

Now, what exactly *is* a library? According to ‘Librarians’ Glossary’ by Harrod (1977:487), a library is (at least) two things. First, it is a “collection of books and other literary material kept for reading, study and consultation”. In other words, it is a *selection* of books. Secondly, a library is “a place, building, room or rooms set apart for the keeping and use of a collection of books”. Thus, the first meaning focuses on the books, whereas the second is concerned with the physical infrastructure. A library is not necessarily and only its books, and vice versa.

The person who puts order in the library is the librarian (Harrod, 1977:486-487):

one who has to care of a library and its contents, selecting the books, documents and non-book materials which comprise its stock, and providing information and loan services to meet the needs of its users.

A definition of book selection is “the process of choosing books for inclusion in a
library with a view to providing a balanced increase to the stock” (129). The librarian needs to catalogue all information in the library, which means to compile a list of documents according to a set of rules so as to enable the consulter to know what items are available, and from the class number, call number, or other means of identification, where they may be found.

Since the library consists of rooms as well as book-holders (e.g., shelves), many possibilities exist to distribute physical books among physical places.

The physical ordering is part of the design of a library, basically a subfield of architecture (Roth, 2011). Buildings, including libraries, can be seen as graph structures, where the nodes are rooms and the connections are corridors and passages. Edwards (2009) extensively surveys many (functional) design aspects for modern libraries. Physical libraries consist of rooms with bookshelves filled with books. The library’s books are distributed over the physical space according to the order, and the catalogue is required to find and locate any book in the physical space. A library user traverses the corridors and the rooms, searching for specific books in the catalogue, or browsing to find interesting books, guided by the physical order in the collection. The logical ordering of the books according to a catalogue is orthogonal to the physical distribution over the library and deals with how to form subgroups of books that “belong together”, for example, based on features such as author’s name, title, subject, size, colour, and so on (Edwards, 2009:106-111; Petroski, 1999:appendix). The physical dimension prevents us from rearranging every day, rendering the order only slowly varying. Shelves generally do not align well with any of the feature-based orderings: almost certainly all books according to one feature will not exactly fill a number of shelves, leaving room for books of a different category. A catalogue of books ordered by some feature(s) is still in need of a reference system that tells you where the book can be found in some list, index or catalogue, may be found in the physical location of a room and a shelf. A simple idea (see Eco's preface to Hoefer, 2005) is to use a simple four-digit code for books such as 32-2-4-13, where the first number specifies the room, the second the wall, the third the shelf, and the fourth the book itself. A formal,
general system is the widely used *Dewey-decimal system* that subdivides the collection by subject area, such as philosophy and religion. When new books come in, shelves get rearranged, and books may get new (sub)classifications. The librarian is responsible for keeping the physical books arranged according to the order.

**Universal libraries**

A *universal library* essentially is a library containing all knowledge, or alternatively, *all possible books*. As White (2008) and Nerdinger (2011b) both extensively describe, the idea of a universal library is an old one. Over two millennia ago, the Ptolemites started the famous library of Alexandria, possibly the first universal library. Universal here means *as complete as currently possible*. In the physical reality, this means that physical copies of all books need to be assembled at the same (geographical) location. Others have envisioned similar libraries, such as Gessner (1516-1565) (*Bibliotheca Universalis*), but also Naudee (1627) and Bacon (1620). Most writers and thinkers about universal libraries have a vision in which *all knowledge is gathered and organised* to the aid of mankind. Eco’s library described in the introductory section was such a library. The physical nature of books and buildings made all of the approaches towards universal libraries in history rather *local*. That is, a library would be the place of knowledge, literally, since it would be an actual *place* where all the knowledge, books, the librarian and the catalogue would reside.

A second type of universal library is the one we find mostly as an idea in fiction. Borges’ (1941) ‘The Library of Babel’ describes the very idea of an imaginary library of *all* possible books. The only requirement is that each book has the same number of pages (410), lines (40/page) and characters (80/line). He also described exactly how the library building itself would look, composed of hexagonal rooms, each connected to two other rooms each holding four walls of five shelves each, each containing 32 books. This library would contain all books known to man, but that is only the beginning. It would also contain all books that were *never* written, and books that are currently *being written*. It would also
contain a book that is exactly like Eco’s *The Name of the Rose*, but in which the library survives the fire. It would simultaneously hold a copy in which the story ends with Eco himself appearing as the fire brigade commander and putting out the fire. *Every book* one can think of would be in the library, in every language thinkable. This also means that a large majority of books would contain just nonsense: there will be many books containing mostly the character ‘a’ and, for each existing novel, there would be copies in the library with any combination of any possible spelling or grammatical error one can think of. The strength of the library is at the same time its weakness: since for any useful or even ‘right’ book, there will be many other books that are useless or books that refute the other books. The library as a whole will not contain any useful information, simply because it contains *any* possible book. In order for a library to be useful, it needs to contain a selection of books, and a catalogue. With reference to the catalogue, Borges (Borges, 1941:116) writes:

> On some shelf in some hexagon (men reasoned) there must exist a book which is the formula and perfect compendium of all the rest: some librarian has gone through it and he is analogous to a god ... Many wandered in search of Him.

An interesting idea: what would be the ‘perfect compendium’ for a universal library? And how to find it? Obviously, there will be books just repeating the line “I am not the perfect compendium”, but also books pointing to some other book saying “the book in hexagon 26232, wall number 4, bottom shelf, 5th book on the right, is the perfect compendium”, and other books will say that that book is lying. In other words: finding the *bookman*, as he is called in the story, is an ill-defined problem. One may wonder what this says about building a catalogue of the gigantic and unstructured Internet, which is rapidly growing into a universal library.

Much has been written about the *size* of a universal library. Goldbloom Bloch (2008) mathematically analyses the implications of the settings of (the physical aspects of) Borges’ library and shows that the number of books is $25 \times 25 \times \ldots \times 25$ (and that 1312000 times), which is a 1 followed by 1834097 zeros. This number of books would exceed, by far, the physical space of the universe and the number
itself would not even fit in one of Borges’ books. In addition, the number of different orderings of those books is even more astronomical. These staggering numbers point to one thing: a universal library is a conceptual idea, and for each practical library, physical or digital, only a selection of books can be kept, managed and ordered: it calls for a smart librarian to order a universal library.

Using the same principles, other libraries can be thought of, and have been described elsewhere, for example consisting of all videos of a particular length, or one with all possible biological blueprints in the form of DNA (Dennet, 2013). Many novel (Internet-based) electronic libraries have appeared, such as Google Books and Wikipedia. The thought tool of a universal library is important: in electronic domains, a tiny computer programme could generate (instances of) all possible books, by systematical and exhaustive enumeration. Although this would take too much time and space to actually compute, in theory it is possible. Universal libraries transform the question ‘what (kind of) book shall I write?’ into ‘which book shall I write?’.

A library metaphor for control over information

I have briefly described three mechanisms through which algorithms can take control over our information: measurement, prediction and manipulation. To better understand what digital algorithms do, I invoke an analog library metaphor, which materialises these notions. Let us materialise any digital information system as a library and the algorithms controlling it as the librarian. A user utilising the system to obtain information would then be materialised as a library member looking for a specific book, or information (contained in several books) on a particular topic. Measurement in this analog setting can be seen as looking at which books the user is viewing, in which order the user browses the shelves, in which order she wanders through the library building, and which information requests she poses to the librarian. Prediction algorithms subsequently take such measurements from many users in order to induce detailed library member profiles. For example, they
can predict that, if someone walks through this corridor, it is likely that he or she is looking for books on architecture, and that, if a member browses through books X and Y, it is likely that book Z will be considered next. Manipulation algorithms then have complete freedom to change, in essence, anything in the library: the selection of books, the order of the library and the physical structure of the library itself. Furthermore, they can do this in the blink of an eye, in real-time, based on prediction models. They can, for example, put the ‘best’ books near the entrance, cluster appropriate books according to the member’s current quest, or even close down or downgrade ‘undesirable’ collections altogether. Considering the web as a library, and a search engine as a librarian, we can now see that the first ten search results are ‘close to the entrance’, that this selection and order is based on a profile and past search results, and that one might pick any link to read and proceed the journey through the digital library, while the vigilant search engine watches closely.

**The physical-universal continuum and libraryness**

Physical and universal libraries form two ends of a continuum, ranging from purely physical to total (universal) libraries without any physical constraints. Digital libraries lie on this continuum, increasingly closer to universal libraries in terms of what size and content, but still influenced by some physical constraints such as server space and computational time. Perhaps the closest example, when it comes to a digital library in the traditional sense, is Google Books, started in 2004 with the aim to scan, digitise, categorise and make publicly available (physical) books from the world’s finest (university) libraries. As Darnton (2009:44) writes: “so many millions that soon it will have constructed a digital mega-library greater than anything ever imagined, except in the fiction of Jorge Luis Borges”.

Physical libraries have physical solutions for storage and ordering, but the activities of a user too have physical aspects, such as searching for a book, walking through the building and manually manipulating the books. When moving
from the physical to the universal, we see that both the library becomes less physical and algorithms become more important. More specifically, when we move from analog to digital libraries, we observe the following trends. As far as storage is concerned, physical places become less important, duplicates are no problem and book content can be generated on the fly at any location (e.g., on the Internet). Without rooms, shelves and corridors, the relatedness between books that influences physical positions in the library is replaced by flexible hyperlink-like structures that can link books in many ways. In turn, the librarian’s activities of selection and ordering are easier since storage and handling costs are minimal. The order of the books can (dynamically) take many forms once physical constraints are gone. Finally, the user’s activities in physical libraries include walking routes for browsing, searching and locating. Once freed from physical constraints, any book can be accessed instantly, now completely determined by the logical order.

The reason why thinking of libraries helps in understanding algorithms is largely due to something Edwards (2009:246-247) calls “libraryness”:

To most people, the word ‘library’ evokes a mental picture of a particular type of building. The picture is both an external image and an internal one. The form of the library upon which personal and public perception is based draws upon four interconnecting mental constructs. There is the geometry of space, the grasp of mass and surface, and the effects of light colour and other optical phenomena and, most importantly, the presence of people (paraphrases from Markus, 1993:11). These together, plus the overriding presence of books, allow function (or what librarians call ‘operational requirements’) to generate distinctive plans and arrangement. It is the four acting together which carry connotations of ‘meaning’. Such meaning is expressed in the architecture of the building, the meeting of function and the celebration of the civic realm.

Furthermore:

To a typical library user, the building has a recognizable plan and image which are rich in cultural meaning. Those who design libraries have a responsibility to convey ‘libraryness’ through the manipulation of form, space and light. ... The library is, therefore, a type of building whose image is already well established in the collective mind of society. The mental picture of ‘libraryness’ is in this sense a sign – a particular type of
shape and volume which signals a particular function. Society reads the built sign and receives the meaning codes.

Edwards’ concept of ‘libraryness’ is something that informs the metaphor as I develop it here. Algorithms live in digital space, but by seeing the libraryness of digital information sources, relevant images from our collective mind are immediately present to aid in interpreting such informational contexts.

‘The days of shelves are over: the Learning Jungle

Some way along our continuum, we find the Learning Jungle, a (conceptual) project by the architects Rients Dijkstra and Jason Hilgefort (2010). Part of the library of the future, it frees books from their shelves and frees them from any formal order by equipping them with radio frequency identification (RFID) chips, to track their location in space. Instead of having a spatial categorisation and fixed, physical places for the books, the books can now be positioned anywhere in the space of the library, and found through radio waves by any ‘reading device’ (e.g., a smartphone). Interestingly, the Learning Jungle somehow keeps the library physical, i.e., it maintains maximal libraryness, but moves the order more towards a digital one where physical location is less important. People can just move books around as they feel, and place books where they think they fit best. In addition, features can be measured, to modify the order further, e.g., covering book use, social influences (like the placing behaviour of many people in the same library space) and context (such as season).

As the Learning Jungle architects, Dijkstra and Hilgefort (2010:69) propose:

This concept also allows for the idea of a ‘least-in-demand’ cellar: the library can move the most popular books to the most frequented spaces, while shifting the books that are less in demand to more secluded rooms. ... Another possibility is that the library can adapt itself to reflect the various media used during the different seasons, for example: summer reading versus winter reading.
In line with our continuum, Dijkstra and Hilgefort discuss the transition from the traditional library to modern, technological ones, and identify dimensions such as “from physical to digital”, “from static to dynamic”, from “pre-organized to self-organized” and “from fixed to flexible”.

Traversing a library: influencing factors

In the Learning Jungle, the order in the library is determined by the combined influences of many other people acting the in the same library. As I have explained before, when moving to digital libraries, algorithms assume the role of librarians and determine this same order. Libraries of information are dynamic because of many influences. Related settings in the physical world include route planners that will plan your travel by car, or the organisation of a supermarket. Both types of organisation of information or goods will vary over time because of external influences (new roads and new products), but also the behaviour of groups of other people (i.e., other visitors of the same library). For example, if traffic is high in some part of the country, routes may be modified to cope with that, and if some type of product in a supermarket is very popular, the structure of the supermarket may change to put that product in a more prominent place. The point is that the structure of any type of information library may change in real-time, instantaneously, and due to many factors that are in or beyond the user’s control. In general, there is no right order, and certainly there very many possible orders. Order arises dynamically, sometimes specifically for each individual user, based on many possible features; not by pre-defined classification schemes.
I will illustrate three types of information that influence the library order; personal data, social data, and world knowledge. All can be seen as a particular context that determines the library’s structure. The following images evoke our sense of libraryness by translating back to a physical library.

People make algorithms, and thus people influence our digital libraries. However, individually people do a lot more. Based on huge amounts of personal information, on social networks, by browsing webpages, by using mobile phones, and so on, algorithms can generate prediction functions that may be utilised for manipulation,

Figure 1. Based on analysis of the user’s past library behaviour by the librarian and the data miners, the books (e.g., the first 10 results of a search query) are selected.

Figure 4. The behaviour in the library of many other people (left) indirectly changes the grouping and ordering of books for the individual on the right.
as previously discussed. Google, for example, brings together its user data from gMail, Google Search, Youtube, etc., to form detailed profiles of users. We call usage of such information customisation, or personalisation. In terms of the library, what can happen is that the library is structured to bring ‘desired’ items closer to the user. For example, Amazon and other booksellers may give you a list of recommendations based on a similarity of purchased items with other ones, or in association with the preferences of other correlated users. Another aspect may be that a company predicts personal traits (for example, through Facebook data) to limit the number of possibilities. Generally, particular use of a digital library is a strong form of feedback: the simple fact that a user traversed a certain link, or not, is already information that can be used to influence the library order. Some software also limits the number of options, e.g., frequently used menu items, based on past use. In addition, a user can often customise a portal or search engine with his interests, demographic information or importance ratios of topics.

In addition to personal influence, how one’s information appears is influenced by social aspects. The fact that many people use the same library brings changes to the library beyond the control of an individual. Thus, in the Learning Jungle; the very structure of the library was determined by the combination of all other people’s actions, moving around books, borrowing books, and so on. In digital libraries, we can observe the same thing. Web search engines work, for the most part, based on how people use it. If more people visit a website, or more people link to a website, it becomes more popular, and the search engine places it earlier in the library order, and this will generate even more visitors. In digital libraries this means that the structure will reflect this collective behaviour over time. This phenomenon is also related to internet selling websites that send you advice on “people who bought X also bought Y”. In this second setting, searching for a particular book may still be relatively simple, but searching for a particular topic may have varying results over time because of other people’s interest and behaviour.
A third aspect influencing digital libraries may be the arrangement of general knowledge. For example, search engines become increasingly better at adjusting search results towards context, which may be task context, time, spatial location, and so on. If I search for ‘jaguar’ while physically located in the jungle of some country, I might be looking for the Wiki page on the animal, whereas if I am in the center of a big UK city, I might be more interested in the car brand. Search results may use such information in more and better ways to adjust our digital libraries online.

Of course, other influences exist. Indirect influences are maybe even more present in the model building phase of (machine learning) algorithms, in the form of (algorithmic) biases mentioned before. Other things now make sense in the library metaphor. For example, the well-known filter-bubble concept described by Pariser (2011) (or narrowcasting, as Vaidhyanathan (2011:83) calls it)) may be seen as a library, where only a small subset of rooms is accessible and all point to each other: this way one stays within the confinement of only a few rooms, just because the librarian notices that you seem to enjoy it.
Gatekeepers and sharing

In the library metaphor, we can take the “gatekeeper” concept (Granka, 2010) quite literally. Search engines are the gatekeepers of all knowledge that lies within their libraries: they determine where you enter, and which rooms you get to see. The consequences for access to knowledge can be significant. Mager (2012) and Bozdag (2013) describe the many forms of bias in search engines. Mager explicitly sees dominant capitalist values becoming embedded in the search results. A major problem with search engines is that it is unclear how search results are produced and why. The original PageRank\(^4\) algorithm of Google was never fully disclosed, and with the recently activated Hummingbird\(^5\) algorithm (which is combination of algorithms), it is even more unclear how the ranking calculation is arrived at. Even if search algorithms were to report all of the decisions and information that determined the search results, it would exceed the capacity for human reading and comprehension, since it depends on vast processes of information gathering, statistical analysis, learning, sampling and filtering. In addition, all influencing factors (such as statistical information about other people, and personalised filters and information) play increasing roles in the search results. Bozdag (2013) shows how different the results are for the same query “Ajax” with three different users, producing results on either a programming language, a mythological figure or a soccer club. Each user received her own library order, without knowing how the order was arrived at.

Reading is a cumulative activity, so it is no surprise that how you see your library changes with every book. However, in the digital world, your library is modified by the algorithms of search engines, other people and many other things. Social reading centers around the idea that one shares what (and when and how) one reads, with other people. Winget (2013) describes four categories: talking to a friend about a book; discussing a book online; formal book clubs; and engaging in discussions in the margin. Many practical systems now support social reading. Sites like GoodReads, and Amazon’s Kindle hardware platform allow users to indirectly ‘communicate’ by making notes in books which can get shared. Kindle
features *popular highlights*, passages in a text that were highlighted by at least two people. Other platforms support *dynamic* books, in which the user can “choose their own adventure” (rendering this a specific universal library of all variants of a particular book). Social reading leaves *electronic earmarks* that influence other people’s books, and libraries.

Social reading nudges people into migrating all their reading efforts to the digital domain, where, as I have discussed, powerful librarians can measure, predict and manipulate. For example, when buying at Amazon, you give away what you buy, but, with e-reading on the Kindle (on IOS), there seems not to be any way to *read* without being tracked and measured. As Alter (2012) suggests: “Your e-book is reading you”. Real librarians, in principle, treat your reading habits as confidential – in fact, not too long ago, reading records were considered confidential – but in digital libraries such privacy is lost. Richards (2012) points to the *perils* of social reading and the hazards for *intellectual privacy*: the idea that for some ideas people need to know that they are not being watched. If all your steps in a digital library are known and even broadcasted through a social network, this may change people’s behaviour to socially accepted norms.

The influence of both gatekeeping and sharing mechanisms materialises into how we (are allowed to) move through our metaphorical libraries, and what we see there.
Some reflections on the librarian

The library metaphor helps us to understand what algorithms do by aggregating all that is algorithmic into the librarian. However, we can do more. An analogy can be made with robots, which are similarly formed by aggregations of algorithms. Some algorithms may interpret visual information, some compute joint positions in the robot arms, and some reason about the plan to navigate to the kitchen. However, the best way to understand the robot is by using the intentional stance (see Dennet, 2013): as a rational being, capable of having beliefs, desires and intentions. The field of human-robot interaction (Jones and Schmidlin, 2011) investigates settings in which humans and robots work together. In such cases, the human has to be able to make predictions, such as that the robot believes that there are two objects on the table, and it may have the intention to pick up the left object. Returning to our librarian, we can say that search engines increasingly try to anticipate our intentions. According to Vanderbilt (2013), in the future of search: “Google is moving from simple data-retrieval to a system that understands how we think and what we want – before we even know we want it”. This is a crucial point: in order to understand the array of Google algorithms, we need to take into account that it has beliefs about us, just like in human-robot situations. In addition, it has beliefs about our intentions and desires, and it has desires of its own, e.g., making profit. For example, it makes sense to say: ‘I get these search results because the search engine knows I want to visit Barcelona, it is four o’clock in the morning, and I have a tendency to buy things in these early hours’. The key point is that we can assume search engines have regular, intuitive capabilities to reason about, and predict, mental aspects of us. We do not have to understand machine learning algorithms in order to understand or predict how search engines work. In other words, we can understand the librarian in any digital library by focusing on its beliefs and other mental attributes. Space restrictions prohibit me from pursuing this direction further, but the idea is to side-step the mathematical algorithmics of machine learning, and directly go to a higher level of the ordering of beliefs,
desires and *intentional librarians*; a level with which we are accustomed.

Another point I want to take slightly further is the notion of *artificially intelligent* librarians. A search engine is used to receiving a few words, and one would give back a list of pointers to individual webpages. The library metaphor exemplifies that which links we obtain, and in which order, are influenced by a plurality of factors. But at least we obtain the links, which resemble the underlying sources of information. Increasingly, search engines try to provide *answers* for queries, without delivering the supporting sources. While I may accept this for “the height of the Eiffel tower”, I would have small reservations for “a good pizza in the neighborhood”, and I would definitely not want to have a search engine answering the question “who is right: the Israeli or the Palestinians”! A long time ago, Vannevar Bush coined the “Memex”, and H.G. Wells wrote about the “world brain”, which were similar to Otlet’s ideas about gathering all world knowledge (Wright, 2014), ideas of a system, where information could be held on microfilm or punch cards, and where scholars could communicate from their desk via telephone with the library, after which they would get the required information almost instantly. Now that we have such a system in our pockets, and call it

![Figure 6. The librarian, as envisioned by H.G. Wells, who would answer questions from outside users and would select the appropriate texts.](image)
‘smartphone’, we should avoid the situation, where the librarian does not let us in the library, but answers all our questions at the door. This prospect becomes more real every day, based on the prediction and manipulation capabilities described in this Chapter, but also due to the increasing use of knowledge by digital librarians. Thus, as Vanderbilt (2013) explains:

These are the pillars of Google’s future of search: the vast knowledge of user behaviour and intent it already has and it compiling every second; the Knowledge Graph, in which strings become things; and Google’s advances in artificial intelligence.

To improve search, you need information; this means that, in order to search in the digital library, the librarian utilises information from that very library. According to Mickiewicz (2011:127): “what search engine companies like Google need to even hope to come up with a new semantic search engine, is a vast collection of books”. Semantics is the new keyword. Indeed, for Vanderbilt (2013):

The work on the semantic graph is to make connections that traditional search might overlook. With the Knowledge Graph, Google has taken a different step towards the future of search: providing answers, not links. This raises the question of authority, long on the mind of Google engineers.

Overall, we can say that digital librarians are getting increasingly smarter and more complex, and this will have many unforeseen implications, but, hopefully, by taking the intentional stance, we may be able to study and understand them better.

Conclusions

This Chapter has introduced a library metaphor to understand the role and power of algorithms in the ordering and visibility of digital information. It has utilised pre-existing knowledge about libraryness in order to offer a novel understanding of algorithms. The study and understanding of new digital situations, ranging from internet shopping and the news industry to massive open online courses (MOOC), can benefit from this approach.
One significant implication for privacy and surveillance is that *Big Brother* has become *Big Librarian*: the biggest threats do not come from oppressive forces of surveillance, but from algorithms acting as friendly librarians who *nudge* and *manipulate*. Among many political questions, one relevant question is simply this: ‘who hired them and who authorised their judgements?’ Privately owned digital librarians tacitly act as gatekeepers for various kinds of information based on their own business models, without the public knowing how and why, and without much public debate on (the desirability and morality of) their role. A related legal question presents itself: ‘who is legally accountable for the behaviour of digital librarians?’ This issue is made even more pressing with algorithms becoming increasingly more *adaptive* and *autonomous*.

Another significant implication concerns the socio-epistemological aspect, when we move from *Big Data* to *Big Knowledge*. Algorithms that use extensive knowledge and pretend to deliver perfect answers to our questions should be approached with absolute skepticism. After all, as I have described, perfect compendia (for universal libraries) are merely fictional ideas, and, for every practical library, however large they may be in terms of terabytes or petabytes, considerable *bias* is at play. Society-wide consequences of *outsourcing* increasing numbers of knowledge-intensive activities to digital librarians are still unknown, but given the potential to shape our social capital of knowledge, we should start to consider them now.

**References**
<http://online.wsj.com/articles/SB10001424052702304870304577490950051438304>


**Notes**
1 See ‘Metaphors we live by’ (Lakoff and Johnson, 1980).
2 See Wright (2014) and Nerdinger (2011b) for Borges’ predecessors writing about similar ideas (e.g., Fechner, Lasswitz, Carroll and Swift).
4 <http://computationalculture.net/article/what_is_in_pagerank>
5 <http://searchengineland.com/google-hummingbird-172816>
6 See Goldman (2011:103) on portalisation.
7 For a comprehensive history see a recent book by Wright (2014).