

A bag but is language nothing of words

(language is nothing but a bag of words)

THIS IS A DRAFT

Michael Murtaugh

In text indexing and other machine reading applications the term "bag of words" is frequently used to underscore how processing algorithms often represent text using a data structure (word histograms or weighted vectors) where the original order of the words in sentence form is stripped away. While "bag of words" might well serve as a cautionary reminder to programmers of the essential violence perpetrated to a text and a call to critically question the efficacy of methods based on subsequent transformations, the expression's use seems in practice more like a badge of pride or a schoolyard taunt that would go: Hey language: you're nothin' but a big BAG-OF-WORDS.

Bag of words

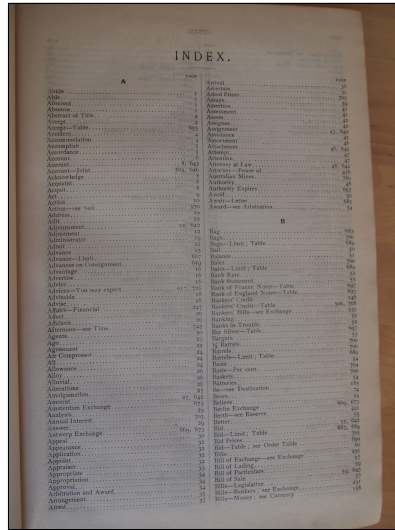
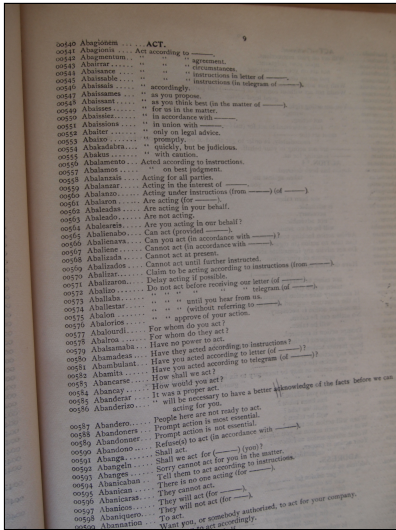
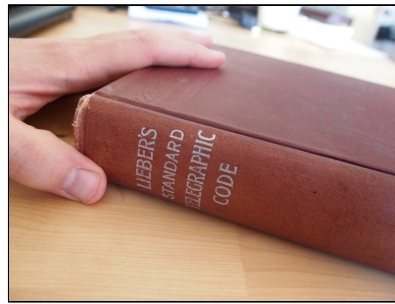
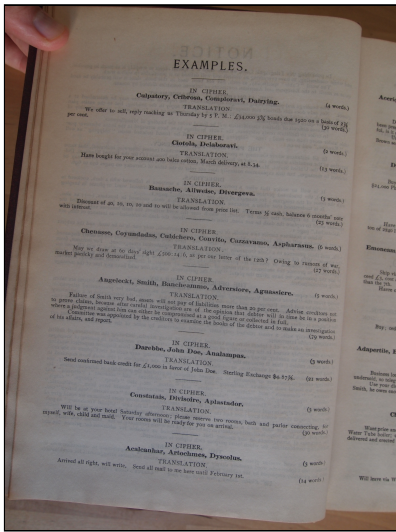
In information retrieval and other so-called *machine-reading* applications (such as text indexing for web search engines) the term "bag of words" is used to underscore how in the course of processing a text the original order of the words in sentence form is stripped away. The resulting representation is then a collection of each unique word used in the text, typically weighted by the number of times the word occurs.

Bag of words, also known as word histograms or weighted term vectors, are a standard part of the data engineer's toolkit. But why such a drastic transformation? The utility of "bag of words" is in how it makes text amenable to code, first in that it's very straightforward to implement the translation from a text document to a bag of words representation. More significantly, this transformation then opens up a wide collection of tools and techniques for further transformation and analysis purposes. For instance, a number of libraries available in the booming field of "data sciences" work with "high dimension" vectors; bag of words is a way to transform a written document into a mathematical vector where each "dimension" corresponds to the (relative) quantity of each unique word. While physically unimaginable and abstract (imagine each of Shakespeare's works as points in a 14 million dimensional space), from a formal mathematical perspective, it's quite a comfortable idea, and many complementary techniques (such as principle component analysis) exist to reduce the resulting complexity.

What's striking about a bag of words representation, given is centrality in so many text retrieval application is its irreversibility. Given a bag of words representation of a text and faced with the task of producing the original text would require in essence the "brain" of a writer to recompose sentences, working with the patience of a devoted cryptogram puzzler to draw from the precise stock of available words. While "bag of words" might well serve as a cautionary reminder to programmers of the essential violence perpetrated to a text and a call to critically question the efficacy of methods based on subsequent transformations, the expressions use seems in practice more like a badge of pride or a schoolyard taunt that would go: Hey language: you're nothing but a big BAG-OF-WORDS. Following this spirit of the term, "bag of words" celebrates a perfunctory step of "breaking" a text into a purer form amenable to computation, to stripping language of its silly redundant repetitions and foolishly contrived stylistic phrasings to reveal a purer inner essence.

Book of words

Lieber's Standard Telegraphic Code, first published in 1896 and republished in various updated editions through the early 1900s, is an example of one of several competing systems of telegraph code books. The idea was for both senders and receivers of telegraph messages to use the books to translate their messages into a sequence of code words which can then be sent for less money as telegraph messages were paid by the word. In the front of the book, a list of examples gives a sampling of how messages like: "Have bought for your account 400 bales of cotton, March delivery, at 8.34" can be conveyed by a telegram with the message "Ciotola, Delaboravi". In each case the reduction of number of transmitted words is highlighted to underscore the efficacy of the method. Like a dictionary or thesaurus, the book is primarily organized around key words, such as *act*, *advice*, *affairs*, *bags*, *bail*, and *bales*, under which exhaustive lists of useful phrases involving the corresponding word are provided in the main pages of the volume. ^[1]



[...] my focus in this chapter is on the inscription technology that grew parasitically alongside the monopolistic pricing strategies of telegraph companies: telegraph code books. Constructed under the bywords “economy,” “secrecy,” and “simplicity,” telegraph code books matched phrases and words with code letters or numbers. The idea was to use a single code word instead of an entire phrase, thus saving money by serving as an information compression technology. Generally economy won out over secrecy, but in specialized cases, secrecy was also important.

[2]

In Katherine Hayles' chapter devoted to telegraph code books she observes how:

The interaction between code and language shows a steady movement away from a human-centric view of code toward a machine-centric view, thus anticipating the development of full-fledged machine codes with the digital computer. [3]

$$X = Y$$

Dick Reckard

Last Revision: 15:27, 22 March 2016

0. Innovation of the same

The PR imagery produced by and around the Mundaneum (disambiguation: the



institution in Mons) often suggests, through a series of 'samenesses', a essential continuity between Otlet's endeavour and Internet-related products and services, in particular Google's. A good example is a scene from the video "From industrial heartland to the Internet age", published by The Mundaneum, 2014 , where the drawers of Mundaneum (disambiguation: Otlet's Utopia) morph into the servers of one of Google's data centres.

This approach is not limited to images: a recurring discourse that shapes some of the exhibitions taking place in the Mundaneum maintains that the dream of the Belgian utopian has been kept alive in the development of internetworked communications, and currently finds its spiritual successor in the products and services of Google. Even though there are many connections and similarities between the two endeavours, one has to acknowledge that Otlet was an internationalist, a socialist, an utopian, that his projects were not profit oriented, and most importantly, that he was living in the temporal and cultural context of modernism at the beginning of the 20th century. The constructed identities and continuities that detach Otlet and the Mundaneum from a specific historical frame, ignore the different scientific, social and political milieus involved. It means that these narratives exclude the discording or disturbing elements that are inevitable when considering such a complex figure in its entirety.

This is not surprising, seeing the parties that are involved in the discourse: these types of instrumental identities and differences suit the rhetorical tone of Silicon Valley. Newly launched IT products, for example, are often described as groundbreaking, innovative and *different from anything seen before*. In other situations, those products could be advertised as *exactly the same* as something else that already exists^[1]. While novelty and difference surprise and amaze, sameness reassures and comforts. For example, Google Glass was marketed as revolutionary and innovative, but when it was attacked for its blatant privacy issues, some defended it as just a camera and a phone joined together. The sameness-difference duo fulfils a clear function: on the one hand, it suggests that technological advancements might alter the way we live dramatically, and we should be ready to give up our old-fashioned ideas about life and culture for the sake of innovation. On the other hand, it proposes we should not be worried about change, and that society has always evolved through disruptions, undoubtedly for the better. For each questionable groundbreaking new invention, there is a previous one with the same ideal, potentially with just as many critics... Great minds think alike, after all. This sort of a-historical attitude pervades technocapitalist milieus, creating a cartoonish view of the past, punctuated by great men and great inventions, a sort of technological variant of Carlyle's *Great Man Theory*. In this view, the Internet becomes the invention of a few father/genius figures, rather than the result of a long and complex interaction of diverging efforts and interests of academics, entrepreneurs and national governments. This instrumental reading of the past is largely consistent with the theoretical ground on which the *Californian Ideology*^[2] is based, in which the conception of history is pervaded by various strains of technological determinism (from Marshall McLuhan to Alvin Toffler^[3]) and capitalist individualism (in generic neoliberal terms, up to the fervent objectivism of Ayn Rand).

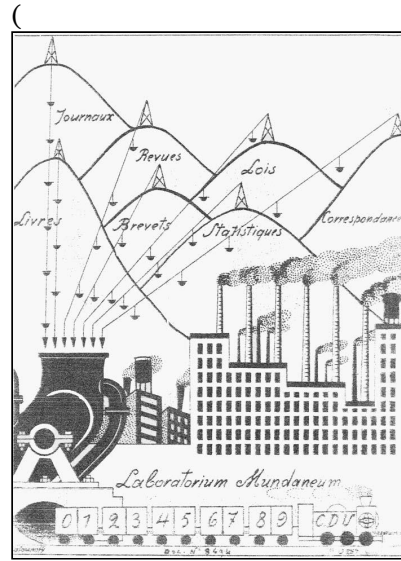
The appropriation of Paul Otlet's figure as Google's grandfather is a historical simplification, and the samenesses in this tale are not without fundament. Many concepts and ideals of documentation theories have reappeared in cybernetics and information theory, and are therefore present in the narrative of many IT corporations, as in Mountain View's case. With the intention of restoring a historical complexity, it might be more interesting to play *exactly the same* game ourselves, rather than try to dispel the advertised continuum of the *Google of paper*. Choosing to focus on other types of analogies in the story, we can maybe contribute a narrative that is more respectful to the complexity of the past, and more telling about the problems of the present.

The Following are three such *comparisons*, which focus on three aspects of continuity between the documentation theories and archival experiments Otlet was involved in, and the cybernetic theories and practices that Google's capitalist enterprise is an exponent of. The First one takes a look at the conditions of workers in information infrastructures, who are fundamental for these systems to work but often forgotten or displaced. Next, an account of the elements of distribution and control that appear both in the idea of a *Reseau Mundaneum*, and in the contemporary functioning of data centres, and the resulting interaction with other types of infrastructures. Finally, there is a brief analysis of the two approaches to the 'organization of world's knowledge', which examines their regimes of truth and the issues that come with them. Hopefully these three short pieces can provide some additional ingredients to adulterate the sterile recipe of the Google-Otlet sameness.

a. Do androids dream of mechanical turks?

In a drawing titled *Laboratorium Mundaneum*, Paul Otlet depicted his project as a massive factory, processing books and other documents into end products, rolled out by a UDC locomotive. In fact, just like a factory, Mundaneum was dependent on the bureaucratic and logistic modes of organization of labour developed for industrial production. Looking at it and at other written and drawn sketches, one might ask: who made up the workforce of these factories?

In his *Traité de Documentation*, Otlet describes extensively the thinking machines and tasks of intellectual work into which the *Fordist chain* of documentation is broken down. In the subsection dedicated to the people who would undertake the work, though, the only role described at length is the *Bibliothécaire*. In a long chapter that describes what education the librarian should follow, which characteristics are required, and so on, he briefly mentions the existence of “Bibliothécaire-adjoints, rédacteurs, copistes, gens de service”^[4]. There seems to be no further description nor depiction of the staff that would write, distribute and search the millions of index cards in order to keep the archive running, an impossible task for the *Bibliothécaire* alone.



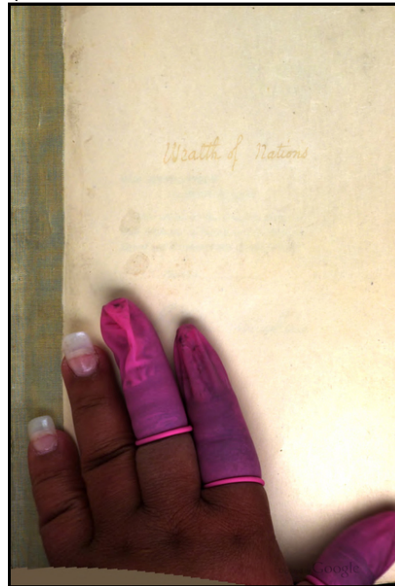
A photograph from around 1930, taken in the *Palais Mondial*, where we see Paul Otlet together with the rest of the equipe, gives us a better answer. In this beautiful group picture, we notice that the workforce that kept the archival machine running was made up of women, but we barely know anything about them. As in telephone switching systems or early software development^[5], gender stereotypes and discrimination led to the appointment of female workers for repetitive tasks that required specific knowledge and precision.



According to the ideal image described in *Traité*, all the tasks of collecting, translating, distributing, should be completely automatic, seemingly without the necessity of human intervention. However, the Mundaneum hired dozens of women to perform these tasks. This human-run version of the system was not considered worth mentioning, as if it was a temporary in-between phase that should be overcome as soon as possible, something that was staining the project with its vulgarity.

Notwithstanding the incredible advancement of information technologies and the automation of innumerable tasks in collection, processing and distributing information, we can observe the same pattern today. All automatic repetitive tasks that *technology* should be able to do for us are still, one way or another, based on human labour. And, unlike the industrial worker who obtained recognition through political movements and struggles, the role of many cognitive workers is still hidden or under-represented. Computational linguistics, neural networks, optical character recognition, all amazing machinic operations are still based on humans performing huge amounts of repetitive intellectual tasks from which software can *learn*, or which software can't do with the same *efficiency*. Automation didn't really free us from labour, it just shifted the where, when and who of labour.^[6] Mechanical turks, content verifiers, annotators of all kinds... The software we use requires a multitude of tasks which are invisible to us, but are still accomplished by humans. Who are they? When possible, work is outsourced to foreign English-speaking countries with lower wages, like India. In the western world it follows the usual pattern: female, lower income, ethnic minorities.

An interesting case of heteromated labour are the so-called Scanops^[7], a set of Google workers who have a different type of badge and are isolated in a section of the Mountain View complex secluded from the rest of the workers through strict access permissions and fixed time schedules. Their work consists of scanning the



pages of printed books for the Google Books database, a task that is still more convenient to do by hand (especially in the case of rare or fragile books). The workers are mostly women and ethnic minorities, and there is no mention of them on the Google Books website or elsewhere; in fact the whole scanning process is kept secret. Even though the secrecy that surrounds this type of labour can be justified by the need to protect trade secrets, it again conceals the human element in machine work. This is even more obvious when compared to other types of human workers in the project, such as designers and programmers, who are celebrated for their creativity and ingenuity.

However, here and there, evidence of the workforce shows up in the result [8].

Whether the tendency to hide the human role is due to the unfulfilled wish for total automation, to avoid the bad publicity of low wages and precarious work, or to keep an aura of mystery around machines, remains unclear, both in the case of Google Books and the *Palais Mondial*. Still, it is reassuring to know that the products hold traces of the work, that even with the progressive removal of human signs in automated processes, the workers' presence never disappears completely. This presence is proof of the materiality of information production, and becomes a sign of the economies and paradigms of efficiency and profitability that are involved.

The computer scientists' view of textual content as "unstructured", be it in a webpage or the OCR scanned pages of a book, reflect a negligence to the processes and labor of writing, editing, design, layout, typesetting, and eventually publishing, collecting and cataloging [9]. (A bag but is language nothing of words)

b. The (data) centre and the periphery

In 2013, while Prime Minister Di Rupo was celebrating the beginning of the second phase of construction of the Saint Ghislain data centre, a few hundred kilometres away a very similar situation started to unroll. In the municipality of Eemsmond, in the Dutch province of Groningen, the local Groningen Sea Ports and NOM development were rumoured to have plans with another code named company, *Saturn*, to build a data centre in the small port of Eemshaven.

A few months later, when it was revealed that Google was behind *Saturn*, Harm Post, director of Groningen Sea Ports, commented: "Ten years ago Eemshaven became the laughing stock of ports and industrial development in the Netherlands, a planning failure of the previous century. And now Google is building a very large data centre here, which is 'pure advertisement' for Eemshaven and the data port."^[10] Further details on tax cuts were not disclosed and, once finished, the data centre will provide at most 150 jobs in the region.



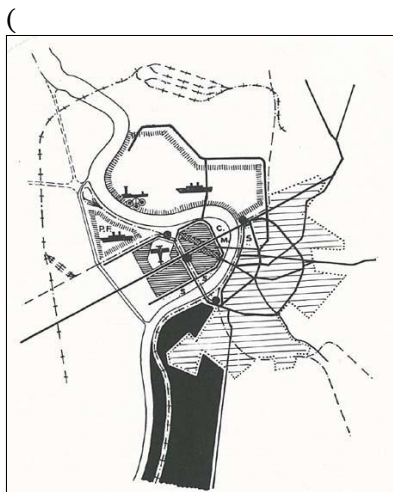
Yet another territory fortunately chosen by Google, just like Mons, but what are the selection criteria? For one thing, data centres need to interact with existing infrastructures and flows of various type. Technically speaking, there are three prerequisites: being near a substantial source of electrical power (the finished installation will consume twice as much as the whole city of Groningen); being near a source of clean water, for the massive cooling demands; being near Internet infrastructure that can assure adequate connectivity. There is also a whole set of non-technical elements, that we can sum up as the social, economical and political *climate*, which proved favourable both in Mons and Eemshaven.

The push behind constructing new sites in new locations, rather expanding existing ones, is partly due to the rapid growth of the importance of *Software as a service*, so-called cloud computing, which is the rental of computational power from a central provider. With the rise of the *SaaS* paradigm the geographical and topological placement of data centres becomes of strategic importance to achieve lower latencies and more stable service. For this reason, in the last 10 years, Google has been pursuing a policy of end-to-end connection between its facilities and user interfaces. This includes buying leftover fibre networks^[11], entering the business of underwater sea cables^[12] and building new data centres, including the ones in Mons and Eemshaven.

The spread of data centres around the world, along the main network cables across continents, represents a new phase in the diagram of the Internet. It should not be confused with the idea of decentralization that was a cornerstone value in the early stages of interconnected networks.^[13] During the rapid development of the Internet and the Web, the new tenets of immediacy, unlimited storage and exponential growth led to the centralization of content in increasingly large server farms. Paradoxically, it is now the growing centralization of all kind of operations in specific buildings, that is fostering their distribution. The tension between centralization and distribution, and the dependence on neighbouring infrastructures as the electrical grid, is not an exclusive feature of contemporary data storage and networking models. Again, similarities emerge from the history of the Mundaneum, illustrating how these issues relate closely to the logistic organization of production first implemented during the industrial revolution, and theorized within modernism.

Centralization was seen by Otlet as the most efficient way to organize content, especially in view of international exchange^[14], which already caused problems related to space back then: the Mundaneum archive counted 16 million entries at its peak, occupying around 150 rooms. The cumbersome footprint, and the growing difficulty to find stable locations for it, concurred to the conviction that the project should be included in the plans of new modernist cities. In the beginning of the 1930s, when the Mundaneum started to lose the support of the Belgian government, Otlet thought of a new site for it as part of a proposed *Cité Mondiale*, which he tried in different locations with different approaches.

Between various attempts, he participated in the competition for the development of the Left Bank in Antwerp. The most famous modernist urbanists of the time were invited to plan the development from scratch. At the time, the left bank was completely vacant. Otlet lobbied for the insertion of a Mundaneum in the plans, stressing how it would create hundreds of jobs for the region. He also flattered the Flemish pride by insisting on how people from Antwerp were more hard working than the ones from Brussels, and how they would finally obtain their deserved recognition, when their city would be elevated to *World City* status.^[15] He partly succeeded in his propaganda; aside from his own proposal, developed in collaboration with Le Corbusier, many other participants included Otlet's Mundaneum as a key facility in their plans. In these proposals, Otlet's archival infrastructure was shown in interaction with the existing city flows such as industrial docks, factories, the railway and the newly constructed stock market.^[16] The modernist utopia of a planned living environment implied that methods similar to those employed for managing the flows of coal and electricity could be used for the organization of culture and knowledge.



The *Traité de Documentation*, published in 1934, includes an extended reflection on a *Universal Network of Documentation*, that would coordinate the transfer of knowledge between different documentation centres such as libraries or the Mundaneum^[17]. In fact the existing Mundaneum would simply be the first node of a wide network bound to expand to the rest of the world, the *Reseau Mundaneum*. The nodes of this network are explicitly described in relation to "post, railways and the press, those three essential organs of modern life which function unremittingly in order to unite men, cities and nations."^[18] In the same period, in letter exchanges with Patrick Geddes and Otto von Neurath, commenting on the potential of heliographies as a way to distribute knowledge, the three imagine the *White Link*, a network to distribute copies throughout a series of Mundaneum nodes^[19]. As a result, the same piece of information would be serially produced and logistically distributed, described as a sort of *moving Mundaneum* idea, facilitated by the railway system^[20]. No wonder future Mundaneums were foreseen to be built next to a train station.

In Otlet's plans for a *Reseau Mundaneum* we can already detect some of the key transformations that reappear in today's data centre scenario. First of all, a drive for centralization, with the accumulation of materials that led to the monumental plans of World Cities. In parallel, the push for international exchange, resulting in a vision of a distribution network. Thirdly, the placement of the hypothetical network nodes along strategic intersections of industrial and logistic infrastructure.

While the plan for Antwerp was in the end rejected in favour of more traditional housing development, 80 years later the legacy of the relation between existing infrastructural flows and logistics of documentation storage is highlighted by the data ports plan in Eemshaven. Since private companies are the privileged actors in these types of projects, the circulation of information increasingly respond to the same tenets that regulate the trade of coal or electricity. The very different welcome that traditional politics reserve for Google data centres is a symptom of a new dimension of power in which information infrastructure plays a vital role. The celebrations and tax cuts that politicians lavish on these projects cannot be explained with 150 jobs or *economic incentives* for a depressed region alone. They also indicate how party politics is increasingly confined to the periphery of other forms of power and therefore struggle to assure themselves a strategic positioning.

c. 025.45UDC; 161.225.22;
004.659GOO:004.021PAG.

The Universal Decimal Classification^[21] system, developed by Paul Otlet and Henri Lafontaine on the basis of the Dewey Decimal Classification system, is still considered one of their most important realizations, as well as a corner-stone in Otlet's overall vision. Its adoption, revision and use until today demonstrate a thoughtful and successful approach to the classification of knowledge.

The UDC differs from Dewey and other bibliographic systems as it has the potential to exceed the function of ordering alone. The complex notation system could classify phrases and thoughts in the same way as it would classify a book, going well beyond the sole function of classification, becoming a real language. One could in fact express whole sentences and statements in UDC format^[22]. The fundamental idea behind it^[23] was that books and documentation could be broken down into their constitutive sentences and boiled down to a set of universal concepts, regulated by the decimal system. This would allow to express objective truths in a numerical language, fostering international exchange beyond translation, making science's work easier by regulating knowledge with numbers. We have to understand the idea in the time it was originally conceived, a time shaped by positivism and the belief in the unhindered potential of science to obtain objective universal knowledge. Today, especially when we take into account the arbitrariness of the decimal structure, it sounds doubtful, if not preposterous.

However, the linguistic-numeric element of UDC, which enables to express fundamental meanings through numbers, plays a key role in the oeuvre of Paul Otlet. In his work we learn that numerical knowledge would be the first step towards a science of combining basic sentences to produce new meaning in a systematic way. When we look at *Monde*, Otlet's second publication from 1935, the continuous reference to multiple algebraic formulas that describe how the world is composed, suggests that we could at one point "solve" these equations, and modify the world accordingly.^[24] As a complementary part to the *Traité de Documentation*, which described the systematic classification of knowledge, *Monde* set the basis for the transformation of this knowledge into new meaning.

Otlet wasn't the first to envision an *algebra of thought*. It has been a recurring *topos* in modern philosophy, under the influence of scientific positivism and in concurrence with the development of mathematics and physics. Even though one could trace it back to Ramon Llull and even earlier forms of combinatorics, the first to consistently undertake this scientific and philosophical challenge was Gottfried Leibniz. The German philosopher and mathematician, a precursor of the field of symbolic logic, which developed later in the 20th century, researched a method that reduced statements to minimum terms of meaning. He investigated a language which "... will be the greatest instrument of reason," for "when there are disputes among persons, we can simply say: Let us calculate, without further ado, and see who is right".^[25] His inquiry was divided in two phases. The first one, analytic, the *characteristica universalis*, was a universal conceptual language to express meanings, of which we only know that it worked with prime numbers. The second one, synthetic, the *calculus ratiocinator*, was the algebra that would allow operations between meanings, of which there is even less evidence. The idea of calculus was clearly related to the infinitesimal calculus, a fundamental development that Leibniz conceived in the field of mathematics, and which Newton concurrently developed and popularized. Even though not much remains of Leibniz's work on his *algebra of thought* it was continued by mathematicians and logicians in the 20th century. Most famously, and curiously enough around the same time Otlet published *Traite* and *Monde*, logician Kurt Godel used the same idea of a translation into prime numbers to demonstrate his incompleteness theorem.^[26] The fact that the *characteristica universalis* only made sense in the fields of logics and mathematics is due to the fundamental problem presented by a mathematical approach to truth beyond logical truth. While this problem was not yet evident at the time, it would emerge in the duality of language and categorization, as it did later with Otlet's UDC.

The relation between organizational and linguistic aspects of knowledge is also one of the open issues at the core of web search, which is, at first sight, less interested in objective truths. At the beginning of the Web, around the mid '90s, two main approaches to online search for information emerged: the web directory and web crawling. Some of the first search engines like Lycos or Yahoo!, started with a combination of the two. The web directory consisted of the human classification of websites into categories, done by an "editor"; crawling in the automatic accumulation of material by following links, with different rudimentary techniques to assess the content of a website. With the exponential growth of web content on the Internet, web directories were soon dropped in favour of the more efficient automatic crawling, which in turn generated at this point so many results that quality has become of key importance. Quality in the sense of the assessment of the webpage content in relation to keywords as well as the sorting of results according to their relevance.

Google's hegemony in the field has mainly been obtained by translating the relevance of a webpage into a numeric quantity according to a formula, the infamous PageRank algorithm. This value is calculated depending on the relational importance of the webpage where the word is placed, based on how many other websites link to that page. The classification part is long gone, and linguistic meaning is also structured along automated functions. What is left is reading the network formation in numerical form, capturing human opinions represented by hyperlinks, i.e. which word links to which webpage, and which webpage is generally more important. In the same way as UDC systematized documents via a notation format, the systematization of relational importance in numerical format brings functionality and efficiency. In this case rather than linguistic the translation is value-based, quantifying network attention independently from meaning. The interaction with the other infamous Google algorithm, AdSense, adds an economic value to the PageRank position. The influence and profit deriving from how high a search result is placed, means that the relevance of a word-website relation in Google search results translates to an actual relevance in reality.

Even though both Otlet and Google say they are tackling the task of *organizing knowledge*, we could posit that from an epistemological point of view the approaches that underlie their respective projects are opposite. UDC is an example of an analytic approach, which acquires new knowledge by breaking down existing knowledge into its components, based on objective truths. Its propositions could be exemplified with the sentences "Logic is a subdivision of Philosophy" or "PageRank is an algorithm, part of the Google search engine". PageRank, on the contrary, is a purely synthetic one, which starts from the form of the network, in principle devoid of intrinsic meaning or truth, and creates a model of the network's relational truths. Its propositions could be exemplified with "Wikipedia is of the utmost relevance" or "The University of District Columbia is the most relevant meaning of the word 'UDC'".

We (and Google) can read the model of reality created by the PageRank algorithm (and all the other algorithms that were added during the years^[27]) in two different ways. It can be considered a device that 'just works' and does not pretend to be true but can give results which are useful in reality, a view we can call *pragmatic*, or instead, we can see this model as a growing and improving construction that aims to coincide with reality, a view we can call *utopian*. It's no coincidence that these two views fit the two stereotypical faces of Google, the idealistic Silicon Valley visionary one, and the cynical corporate capitalist one.

From our perspective, it is of relative importance which of the two sides we believe in. The key issue remains that such a structure has become so influential that it produces its own effects on reality, that its algorithmic truths are more and more considered as objective truths. While the utility and importance of a search engine like Google are out of the question, it is necessary to be alert about such concentrations of

power. Especially if they are only controlled by a corporation, which, beyond mottoes and utopias, has by definition the single duty of to make profits and obey its stakeholders.

1. A good account of such phenomenon is described by David Columbia. <http://www.uncomputing.org/?p=221>
2. As described in the classic text looking at the ideological ground of Silicon Valley culture. <http://www.hrc.wmin.ac.uk/theory-californianideology-main.html>
3. For an account of Toffler's determinism, see <http://www.ukm.my/ijit/IJIT%20Vol%201%202012/7wan%20fariza.pdf> .
4. Otlet, Paul. *Traité de documentation: le livre sur le livre, théorie et pratique*. Editions Mundaneum, 1934: 393-394.
5. <http://gender.stanford.edu/news/2011/researcher-reveals-how-%E2%80%9Ccomputer-geeks%E2%80%9D-replaced-%E2%80%9Ccomputergirls%E2%80%9D>
6. This process has been named “heteromation”, for a more thorough analysis see: Ekbia, Hamid, and Bonnie Nardi. “Heteromation and Its (dis)contents: The Invisible Division of Labor between Humans and Machines.” *First Monday* 19, no. 6 (May 23, 2014). <http://firstmonday.org/ojs/index.php/fm/article/view/5331>.
7. The name scanops was first introduced by artist Andrew Norman Wilson when he found out about this category of workers during his artistic residency at Google in Mountain View. See <http://www.andrewnormanwilson.com/WorkersGoogleplex.html> .
8. As collected by Krissy Wilson on her <http://theartofgooglebooks.tumblr.com> .
9. <http://informaticsobservatory.info/2015/10/27/google-books-fair-use-or-anti-democratic-preemption/#more-279>
10. <http://www.rtvnoord.nl/nieuws/139016/Keerpunt-in-de-geschiedenis-van-de-Eemshaven> .
11. <http://www.cnet.com/news/google-wants-dark-fiber/> .
12. <http://spectrum.ieee.org/tech-talk/telecom/internet/google-new-brazil-us-internet-cable> .
13. See Baran, Paul. “On Distributed Communications.” *Product Page*, 1964. http://www.rand.org/pubs/research_memoranda/RM3420.html .
14. Pierce, Thomas. *Mettre des pierres autour des idées*. Paul Otlet, de Cité Mondiale en de modernistische stedenbouw in de jaren 1930. PhD dissertation, KULeuven, 2007: 34.
15. Ibid: 94-95.
16. Ibid: 113-117.
17. Otlet, Paul. *Traité de documentation: le livre sur le livre, théorie et pratique*. Editions Mundaneum, 1934.
18. Otlet, Paul. *Les Communications MUNDANEUM, Documentatio Universalis*, doc nr. 8438
19. Van Acker, Wouter. “Internationalist Utopias of Visual Education: The Graphic and Scenographic Transformation of the Universal Encyclopaedia in the Work of Paul Otlet, Patrick Geddes, and Otto Neurath.” *Perspectives on Science* 19, no. 1 (January 19, 2011): 68-69.
20. Ibid: 66.
21. The Decimal part in the name means that any records can be further subdivided by tenths, virtually infinitely, according to an evolving scheme of depth and specialization. For example, 1 is “Philosophy”, 16 is “Logic”, 161 is “Fundamentals of Logic”, 161.2 is “Statements”, 161.22 is “Type of Statements”, 161.225 is “Real and ideal judgements”, 161.225.2 is “Ideal Judgements” and 161.225.22 is “Statements on equality, similarity and dissimilarity”.
22. “The UDC and FID: A Historical Perspective.” *The Library Quarterly* 37, no. 3 (July 1, 1967): 268-270.
23. TEMP: described in french by the word *depouillement*,
24. Otlet, Paul. *Monde, essai d'universalisme: connaissance du monde, sentiment du monde, action organisée et plan du monde*. Editions Mundaneum, 1935: XXI-XXII.
25. Leibniz, Gottfried Wilhelm, *The Art of Discovery* 1685, Wiener: 51.
26. https://en.wikipedia.org/wiki/G%C3%B6del_numbering
27. A fascinating list of all the algorithmic components of Google search is at <https://moz.com/google-algorithm-change> .