



HAL
open science

Search engine literacy

Olivier Le Deuff

► **To cite this version:**

Olivier Le Deuff. Search engine literacy. Information Literacy in the Workplace. ECIL 2017, Sep 2017, Saint-Malo, France. sic_01796309

HAL Id: sic_01796309

https://archivesic.ccsd.cnrs.fr/sic_01796309

Submitted on 19 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Search engine literacy

Olivier Le Deuff

University of Bordeaux Montaigne, MICA, Bordeaux, France,
oledeuff@gmail.com

Abstract. The aim of this article is to show that search engine literacy (SEL) can be connected as a transliteracy to digital literacy and information literacy. SEL allows a better understanding of the relationship between information retrieval and technical systems. We show here that teaching Search Engine Optimization (SEO) can be a part of SEL.

Keywords: literacy, transliteracy, search engines, search engine literacy, search engines, google.

1 Introduction

If SEO (search engine optimization) is well known, the aim of his article is to show that people's search engine literacy is often very poor. As a professor, I ask my freshmen students every year if they can explain how Google works and how it ranks websites. Not one of my student can answer this question. This is particularly disturbing because my students want to work with information or communication systems. Moreover it demonstrates the failure of their previous education, which should have given them the key to understanding how the main search engines operate.

Many studies have shown that students have great confidence in the results given by Google [1] [2], but they do not develop any critical thinking about its ranking methods.

Moreover, the younger generation is not the only one which does not understand how Google works and how google makes money. In a study of seniors and digital health literacy [3], we have shown that Google was the first tool used to access information on the web (92%). But the design of results is not totally understood by these older users, who very often cannot make the distinction between 'natural search results' and ads [4]. At the end of this study, we developed an application to help users to evaluate their level of digital health literacy¹. The results show that the level of health literacy was better than the level of digital literacy because more than 70% of the respondents did not know how Google works.

¹ <http://www.megatopie.info/testor/enquete.html?e=19>

The aim of this article is to consider search engine literacy (SEL^o) as an essential part of transliteracy[5]. In fact, search engines like Google are a new kind of media, and major players in information retrieval. Search engine literacy offers the possibility to study the operating mode of Google and to study algorithms from a different perspective from mathematical or computer science approaches.

This literacy cannot be only grounded on good practices in the use of Google, like the demonstration of advanced research based on a simple search literacy [6] (Wilson et al., 2016). The project of Search engine literacy is more ambitious and calls for a new pedagogy of Google. Later in this article we report on some results of an experiment with our students during a SEO course.

This literacy is a good starting point for explaining the history of the web and the development of a digital and information culture. This is a good method for understanding data literacy and the methods of data treatment. Search engine literacy makes the link between information literacy and data literacy too.

Methodology

This article wants to show the importance of SEL and the relationship with information literacy and culture of information. We want to show here the basis of SEL. This is an exploratory project based on many years of teaching SEO with students in a French i-school. We show here also a project of a French teacher-librarian Florian Reynaud who develops à specific tool to explain search engines.

2 A new literacy?

2.1 The search engine tradition

SEL is not a totally new literacy. We need to consider older tools for information retrieval before the success of search engines like Google. We think about bibliographic systems, index systems, card systems (like the *Repertoire bibliographique universelle* of Paul Otlet) and other incredible tools like the machine of Emanuel Goldberg [7]. Our ways of searching for information have always been connected with technologies. We always need to imagine sophisticated systems to retrieve information and to satisfy the extent of our information needs. Among these original devices, we think of Jean Hautfuney and the creation of the index of the *Speculum Historiale* written by Jean de Beauvais [8]. That was the beginning of the tradition of indexes and tools for information retrieval (*statim invenire*). After bibliographies and index cards, a new kind of index was created by the Belgian Paul Otlet: the index of all printed documents [9]. It was the ‘Google of paper’². We are not in the first age of information, and search engines are not the first generation of ‘knowledge machines’ [6]. In our age of information, the novelty lies in the computational nature of a search engine's features, which operate on automatic indexing and automated language processing. Nevertheless, it is not impossible to

develop historical knowledge of the operating mode of search engines. If learning the information search process requires knowledge of the basic uses of search engines, an ambitious literacy should, rather, involve understanding how they work. That means understanding the context that led to the emergence of Google, in particular the period when leaders like Alta Vista saw a degradation in the quality of their results. Web metadata had been polluted by spam keywords. Pornographic keywords and the name of Pamela Anderson³ were used by webmasters to increase their traffic.

The logic is therefore to find a way to grant credit to a resource listed by its URL. The choice made by Sergey Brin and Larry Page was to base their classification algorithm on the importance of references. They chose to build a citation index of the web. This logic was not new, it comes mainly from the work of scientometrics especially that of Garfield, Pritchard and Nalimov. Google belongs to a documentary tradition based on citation logic and on bibliographic methods originally focused on the study of bibliographic references from journals. However the corpus is different since the web is no longer a web of documents for researchers, but a collection of disparate websites. The simplicity and clean interface of Google conceal a complexity that is not understood by users.

2.2 Search engine literacy and the other literacies

SEL is not a totally new literacy and even less is it disconnected from other literacies that involve information skills on the web. SEL can be considered as a transliteracy that mobilises information skills, digital skills and communication skills.

It is therefore not a separated literacy but rather a literacy that is potentially part of training plans either on the side of information literacy or on the side of digital literacy. It can thus be perfectly described as a transliteracy that makes the link between the two literacies. SEL is also a communication and media literacy. Indeed, we should think of the search engine as a mediation tool that influences content. The evolution of Google makes it a medium in its own right. We find ourselves here aligning with discourses around new media literacies [10].

The challenge is clearly to understand the articulations between content and media.

Search engine literacy allows the link with a citizen conception of information literacy [11], which is precisely a culture of information [12]. This cultural approach implies going beyond good practices to develop a wider and more sustainable knowledge by teaching information and digital concepts. This means better ability to understand the tools used and their foundations. It is also a better cognition and recognition of the technique as part of culture [13][14]. This implies, therefore, the need to learn the history of the web and the evolution of search engines.

³ Les meta-tags hors la loi ? < <http://www.sam-mag.com/archives/metatags.htm>>

3 SEO as a strategy of SEL

The challenge is to consider a didactic perspective in teaching search engines, to overcome the “black box” effect, and especially to be able to analyse the results or even criticise them. Indeed, the algorithmic system of search engines relies on computational logic, rather than on logic of analysis. This explains why some research results are disturbed by the high ranking of bad websites with fake news and which are far from the historical truth. For example, a Google search about the historians of the shoah puts forward some negationist websites. This means information literacy must evolve to integrate more and more digital skills. Search engines hide a complexity with a simple interface but the aim is to open the ‘black box’ following the advice of Johanna Drucker about the necessity to examine the framework:

A frame can extend, intensify, connect, embed, juxtapose, or otherwise modify another frame and perception. The terminology is spatial and dynamic. It describes cognitive processes, not simple actions of an autonomous user, but codependent relations of user and system. In invoking frame analysis as part of the diagrammatic model of interpretation, we have moved from a traditional discussion of graphical formats as elements of a mise en page to a sense that we are involved with a mise en scene or système. This puts us on the threshold of interface and a theory of constructivist processes that constitute the interface as a site of such cognitive relations. Interface is not a thing, but a zone of affordances organized to support and provoke activities and behaviors probabilistically. [15, p.7]

3.1 Learning SEO as a core skill for Informationschools

One of the ways to better understand the way search engines work, and especially their ranking methods, is to look at it from an SEO perspective. Students can better understand the key elements used by google (title markup, domain names, keywords), the page rank and the choices of weighting used in the algorithm.

The observations were made during a course of SEO in a small I-school in France in Bordeaux in DUT Information numérique dans les organisations ⁴(digital information in organizations) where we train students for two years. SEO is just a course, which is connected with other courses like metadata history and taxonomies, and markup languages like html and xml. Search engine literacy needs to be connected to many other elements if we want to place it as core content in an I-school curriculum. We can list some essential content:

- The constitution of an index
- History of indexes

⁴ <http://www.infonumbordeaux.fr/>

- Metadata history, systems and languages
- Human indexing and automatic indexing
- Crawl methods and bots
- Web directories versus Search engines
- Distinction between natural indexing and advertising indexing
- Scientometrics and page rank
- Algorithm and weighting of criteria
- Keyword optimization
- Communication strategies and SEO
- Ethics and search engines
- Personal Data and filter bubbles.

For three years we have experimented with having an SEO contest between groups of students as part of the course. The aim is for them to imagine they are content producers. It is also to apply the SEO methods they have learnt about, to try to place their sites or contents top of the hit list on Google. Each year, the choice is based on a request where there are almost no results so that the student productions can best be placed in the first results. The aim of the approach is certainly to be trained in SEO, but the objective is also to pursue the process of training in information retrieval by placing the students on the side of content producers. SEL makes it possible to consider both sides of the information, at the level of the search and also at the level of its production. It makes it easier for students to judge bias in search engines. The principle of motivating students through a logic of challenge or competition also shows that even if search engines avoid too much optimization, search engines' ranking methods differ from a human analysis as regards checking content quality. Therefore, dual training in information retrieval methods and SEO strategies makes it possible to go further in detecting content which is rubbish". In French, we use 'évaluation de l'information' to describe the analysis of bad contents and the work which allows one to distinguish good contents and relevant websites. If this type of course seems essential for an I-School education, it is advisable to extend it earlier on. Without necessarily training in SEO, we could teach how search engines work, especially Google, in an educational perspective.

3.2 Search engine pedagogy

Several hypotheses can be envisaged, and are also tested by French colleagues who are teacher-librarians (called in French *professeurs-documentaliste*) with their students, to show them how research results are linked to the constitution of an index⁵. Tools such as Google custom search can be used to create thematic engines, but you

⁵ ID Base. Fiches pédagogiques à l'usage des professeurs documentalistes.
<<http://idbase.esmeree.fr/>>

can also rely on specific engines. A *webfinder simulation*⁶ is proposed by *professeur-documentaliste* Florian Reynaud in order to be able to use it in class with students. The device makes it possible to understand the steps of classification and how results are ranked. The methods of automatic processing are also explained. It enables understanding of the criteria for ranking information according to the principles of relevance (number of occurrences in connexion to the request), popularity (number of visits), and notoriety (number of links). This also helps learners to understand what an algorithm is with a weighting system.

WebFinder
moteur de recherche

Rechercher...

Résultats de la recherche :

1. Digestion humaine - Wikipédia
fr.wikipedia.org/wiki/Digestion_humaine
Cet article traite du développement, de l'organisation et du fonctionnement du système digestif de l'homme. Sommaire 1 Développement embryologique du tube digestif 1.1 Généralités 1.2 Organisation 1.2.1 Œsophage 1.2.2 Estomac 1.2.3 Pancréas ...
2. Alimentation - digestion - absorption - métabolisme - excréation
pitt.chésa.alice.fr/virtufo/mangee.htm
L'alimentation ou prise alimentaire ou encore ingestion est la première étape de la fonction digestive au sens large. Les aliments liquides sont avalés, les aliments solides sont mastiqués (appareil masticateur comprenant les dents) et subissent un début de digestion (modification chimique) par les ...
3. Comment digérons-nous les aliments ? - EurekaSanté par VIDAL
eureka.sante.vidal.fr/autre/ftion/corpo-aliments/digestion
Comment digérons-nous les aliments ? Mis à jour : Lundi 30 Novembre 2009 42 Sommaire Comment digérons-nous les aliments ? La digestion commence dans la bouche La digestion continue dans l'estomac Le rôle de l'intestin grêle dans la digestion Le rôle du gros intestin dans la digestion De la bouche a...
4. Leçon sur notre système digestif
www.larousse.net/fr/sciences/letra_humain/lecon_sur_notre_notre_système_digestif_(leçon) DEFINITIONS ET LOCALISATIONS La langue | Organe du goût. Les dents | Éléments implantés dans les mâchoires servant à couper et broyer les aliments. Le pharynx : Partie du tube digestif située entre la bouche et l'œsophage. L'œsophage : Partie du tube digestif situ...

Observe les résultats (dans la colonne de gauche).

D'après toi, et en observant la base de données, sur quels deux critères sont ordonnés les 10 résultats ?

Quel critère est le plus important ?

La base de données :

Tu peux trier chaque colonne en cliquant sur un titre de colonne. Pour trier deux ou trois colonnes, maintiens la touche SHIFT enfoncée et clique sur la nouvelle colonne pour trier.

Titre de la page web (-> titre >)	Adresse URL *	Texte	Nombre de liens, sur le s Web, vers cette page	Nombre de visites sur cette page	Score pour les visites	Nombre de mots du lemme "digestion" dans le texte	Score pour les mots	Page Rank
Comment digérons-nous les aliments ? - EurekaSanté par VIDAL	http://eureka.sant...	Comment digérons-nous les aliments ? Mis à jour : Lundi 30 Novembre 2009 42 Sommaire Comment digérons-nous les aliments ? La digestion commence dans la bouche La digestion continue dans l'estomac Le r...	21	24520	8	24	6	14

Webfinder, a tool to understand search engines methods of ranking

SEL pedagogy is an important step for information literacy which can show the complexity of the devices and goes beyond the simple satisfaction of an information need.

Conclusion

The description of a culture of information for citizens requires a better knowledge of the information retrieval processes and the tools used. SEL can't be considered as a specific literacy. It's a kind of transliteracy and it needs to be connected with information literacy. SEL show the evolution of the methods of teaching information literacy with the will to explain how search engines work and not only the best practices to use them.

⁶ <https://iddocs.fr/webfinder/>

Better understanding of search engines, especially through examination of the criteria used for ranking websites, is an important step in the democratization of the search engine and especially Google. It also gives users the means to know when to use Google and when to favour other search methods. Google does not want to publish their algorithm because the firm pretends to protect the truth and to avoid too much SEO optimization. This position shows, too, the will to protect a business model. The democratization of Google appears difficult in this model if the users are not well trained and cannot understand the strength and the weakness of such a model. Training students can be the way to imagine that other kind of search engines could be invented in the next years. Creating an open index [16] of the web could allow alternate models.

References

1. Hargittai, E., Fullerton, L., Menchen-Trevino, E., Yates-Thomas, K.. Trust Online: Young Adults' Evaluation of Web Content. *International Journal of Communication*, vol.4, p. 468–494. (2010)°
2. Simonnot, B. . De l'usage des moteurs de recherche par les étudiants. In Simonnot, B., Gallezot, G.ir.). *L'entonnoir. Google sous la loupe des sciences de l'information et de la communication*. Caen: C&F Editions. p. 31–57 (2009)°
3. Vigouroux-Zugasti, E, Le Deuff, O. Le senior connecté. in Saleh, I., et ali (dir.). *H2PTM'15*, ISTE Editions, p. 174-184 (2015)
4. Le Deuff, O.. La littératie digitale de santé, un domaine en émergence in Agostinelli, Serge, Koulayan, Nicole, *Les écosystèmes numériques : Intelligence collective, développement durable, interculturalité, transfert de connaissances*, Presses des Mines, 2016, p.349-369 (2016)
5. Thomas,S et al. 'Transliteracy: Crossing divides.' *First Monday*, Volume 12 Number 12 - 3 December (2007), :<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2060/1908>
6. Wilson, Max L. and Ye, Chaoyu and Twidale, Michael B. and Grasse, Hannah and Rosenthal, Jacob and McKittrick, Max (2016) Search literacy: learning to search to learn. In: *Second International Workshop on Search as Learning (SAL 2016)*, 21st July, Pisa, Italy (2016)
7. Buckland, Michael Keeble. *Emanuel Goldberg and His Knowledge Machine*. Greenwood Publishing Group, (2006).
8. Paulmier-Foucart, M. « Ut diligens et intelligens lector requirat : une banque de données du Speculum maius de Vincent de Beauvais », *De la plume d'oie à l'ordinateur : études de philologie et de linguistique offertes à Hélène Naïs*, (Verbum, numéro spécial, Université de Nancy), Nancy, p. 469-480. (1985)°
9. Olivier Le Deuff, O « Utopies documentaires : de l'indexation des connaissances à l'indexation des existences », *Communication et organisation*, 48 | 2, 93-106. (2015)
10. Jenkins, H. *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. MacArthur Foundation, (2008).
11. Owens, M. State. *Government and Libraries*. *Library Journal*, (1976).

12. Le Deuff, O « Penser la conception citoyenne de la culture de l'information », Les Cahiers du Numérique. « La culture informationnelle », vol.5, n° 3, p. 39-49 (2009°)
13. Stiegler, Bernard. La technique et le temps. 1., Paris: Galilée : Cité des sciences et de l'industrie (1994).
14. Simondon, Gilbert. Du mode d'existence des objets techniques. Paris: Editions Aubier,(2012).
15. Drucker, Johanna. "Humanities Approaches to Interface Theory." Culture Machine 12 (2011).
<http://svr91.edns1.com/~culturem/index.php/cm/article/download/434/462>.
16. Lewandowski, Dirk. "Why We Need an Independent Index of the Web." arXiv:1405.2212 [Cs], May 9, (2014°). <http://arxiv.org/abs/1405.2212>.

Acknowledgment

I want to thank Sheila Webber for reviewing this article.
Thanks to Florian Reynaud for the screenshot of his *Webfinder*.