

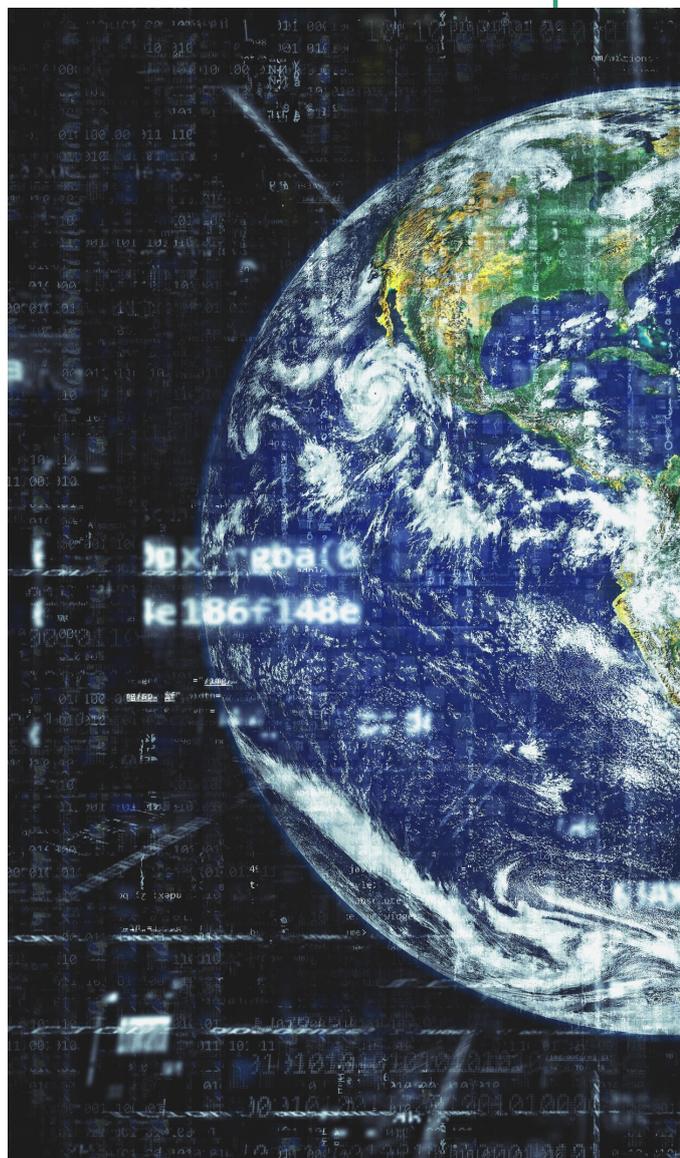
OCTOBRE 2020

# LES MOTEURS & MÉTAMOTEURS DE RECHERCHE

Etude menée par  
Informatique CDC

*Rapport réalisé par la communauté  
Do IT Green d'Informatique CDC*

**CDC** INFORMATIQUE



### PETIT PRÉCIS

## de recherche sur internet

## CHIFFRES MARQUANTS

**40 000**

**Enfants travaillent dans des mines**, de cobalt notamment, un composant essentiel des batteries de nos produits numériques, dans le sud du Congo. (Source : Amnesty International, «Voilà pourquoi on meurt », Rapport, 2015)

**3.8 %** des émissions de GES

**Impact du numérique mondial** (Source : GreenIT.fr, Empreinte environnementale du numérique mondial, 2019)

**53.6 millions**

**de tonnes de DEEE au niveau mondial** (1.4 millions pour la France, 21kg par personne, 54% officiellement collecté et recyclé) en 2019 (Source : United Nations University, The Global E-waste Monitor 2020 : Quantities, flows and the circular economy potential, 2020)

**Google.** Probablement un des mots les plus connus et utilisés au monde à l'heure actuelle. Qui n'a jamais entendu parler du géant américain du numérique ? La firme californienne fait partie intégrante de nos vies à travers ses nombreux services numériques et notamment son service historique : le moteur de recherche. Plus de 90% des recherches internet mondiales utilisent désormais Google [1] ! Cet outil d'accès à l'information sur internet nous paraît très familier, mais sait-on seulement comment il fonctionne ? À l'heure où les impacts environnementaux et sociaux du numérique apparaissent au grand jour en parallèle de la prise de conscience de l'urgence écologique, **il semble primordial de réfléchir à nos usages numériques.**

Le moteur de recherche internet, en tant qu'outil d'usage quotidien d'une grande majorité - voire de l'ensemble - d'entre nous, semble être une porte d'entrée intéressante. Le numérique n'est pas une technologie détachée du monde réel et sans impact sur celui-ci, mais au contraire repose sur des choix techniques, d'architecture et de services impactant. Une fois éclairé sur le fonctionnement de cet outil, et de la galaxie d'acteurs existant dans l'ombre de Google, chacun sera en mesure de faire un choix, en pleine conscience de son mode d'utilisation. Dans une logique de transparence quant au périmètre de l'article, il est précisé que seuls les moteurs de recherche internet généralistes seront abordés, laissant de côté les cas plus spécifiques des moteurs spécialistes (recherches d'emploi, bibliographie scientifique...) et moteurs d'entreprise.

[1] DUFFEZ Olivier. Parts de marché des moteurs de recherche (France, USA, monde). Web Rank Info [En ligne]. 2020. Page consultée le 18/08/2020. Accessible au : <https://www.webrankinfo.com/dossiers/etudes/parts-marche-moteurs>

# QU'EST-CE QU'UN Moteur de recherche ?

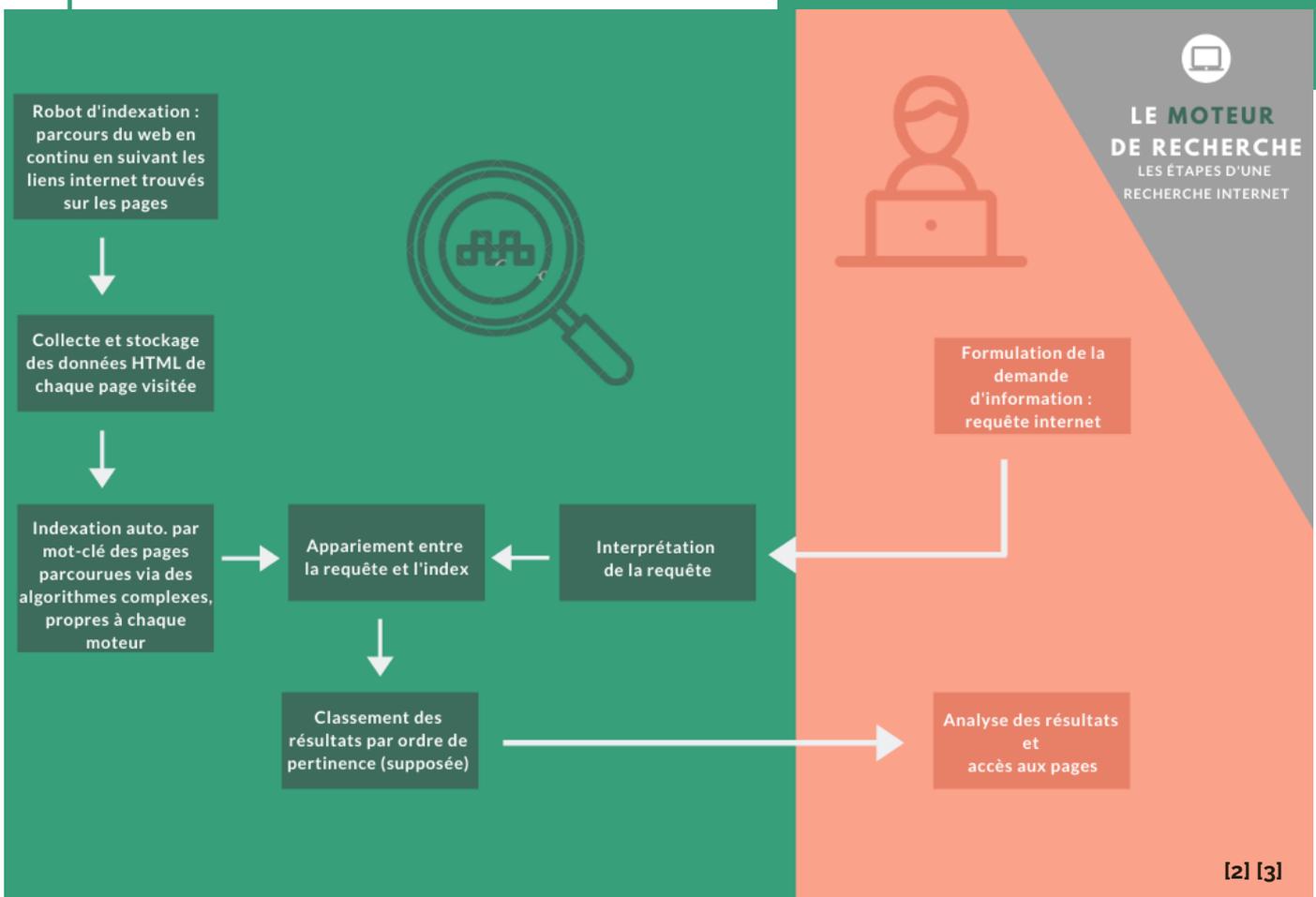
Tout d'abord, il faut différencier les moteurs de recherche des navigateurs Internet. Le navigateur Internet est le logiciel installé sur votre ordinateur (par exemple, Internet Explorer, Mozilla Firefox, Google Chrome, Safari...) que vous utilisez pour accéder aux sites internet, dont les moteurs de recherche sont un type particulier.

Le moteur de recherche, quant à lui, est une application web qui permet de rechercher en ligne des ressources, des contenus, des documents etc., à partir de mots clés. Il suffit de renseigner les expressions qui forment la requête et le moteur de recherche détecte automatiquement les ressources correspondant à la recherche. Les résultats apparaissent organisés selon une logique propre à chaque moteur.

[2] LERAT Victor. Moteur de recherche : qu'est ce qui se cache sous le capot ? Victor-lerat.fr. 2013. Accessible au : <https://www.victor-lerat.fr/fonctionnement-moteur-recherche.html>

[3] Hardwick Joshua. How Do Search Engines Work and Why Should You Care ? AhrefsBlog.com. 2019. Accessible au : <https://ahrefs.com/blog/how-do-search-engines-work/>

## LES ÉTAPES CONSTITUANT UNE RECHERCHE INTERNET



Ce serait toutefois une erreur que de rester naïf une fois devant cette liste de résultats s'affichant, presque comme par magie, en moins d'une seconde, sous nos yeux. Cette liste ne reflète pas la vérité absolue, compilant de manière parfaite et par ordre décroissant de pertinence les résultats à la requête formulée par l'utilisateur.

Tout d'abord, **les résultats varient d'un moteur à l'autre du fait d'algorithmes différents**. Des controverses émergent d'ailleurs régulièrement concernant d'éventuels biais (sexistes, racistes...) nouvellement découverts sur des moteurs de recherche malgré les évolutions constantes de leurs algorithmes. **Les résultats fournis ne sont donc pas forcément complètement neutres**.

Deuxièmement, **la concordance maximale entre attentes de l'utilisateur et résultats fournis peut avoir une contrepartie : les données personnelles**.

Un moteur tel que Google se base en effet sur les méta-données de l'utilisateur (pays, ville, date, heure, informations personnelles contenues dans le profil Google/Facebook connecté sur le PC/téléphone...) pour renvoyer les résultats correspondant le mieux à la fois à la requête mais également à l'utilisateur la formulant.

Google utilise également l'historique de recherche de l'utilisateur[4].

Cette pratique induit donc **un biais de confirmation**, l'utilisateur se retrouvant en moyenne plus souvent face à des résultats proches de ses convictions ou activités habituelles. Ce phénomène est désormais connu sous le terme de « **filter bubble** »[5].

L'accès aux informations sur le comportement d'un utilisateur en ligne (sites visités, contenus visionnés...) permet également aux sites internet de mettre en place un système de « ciblage comportemental »[6].

Concrètement cela signifie **qu'à partir des informations collectées notamment via des cookies sur le comportement en ligne d'une personne, le site lui présentera des publicités correspondant le mieux possible à ces activités**.

Les grands moteurs actuels se rémunèrent ainsi principalement via la publicité, les annonceurs payant les moteurs pour que leurs publicités apparaissent aux utilisateurs lors de leurs recherches.



En parallèle des limites des résultats proposés par les moteurs de recherche, il est bon d'avoir également conscience que les résultats fournis ne « tombent pas du ciel » directement sur notre écran. **Chaque recherche présente un impact environnemental et social trop souvent oublié**. Ainsi, selon l'ADEME (Agence de la transition écologique), **une donnée numérique** (requête web par exemple) **parcourt en moyenne près de 15 000 kilomètres**[7].

Ce **parcours se fait sur l'infrastructure physique du réseau internet mondial**, composée des terminaux de l'utilisateur (ordinateur ou smartphone pour l'envoi de la requête, box fournissant l'accès à internet...), des datacenters stockant les informations et des câbles réseaux faisant le lien entre utilisateurs et datacenters. **Chaque élément de cette infrastructure nécessite d'être produit, alimenté en énergie lors de son fonctionnement et traité une fois en fin de vie**. Chacune de ces étapes engendrant de nombreux impacts environnementaux et sociaux. A titre d'exemple, le **smartphone 5 pouces** utilisé pour envoyer la requête web aura nécessité **l'extraction d'environ 183kg de matières pour sa production**[8].

[4] How Search algorithms work. Google.com. Page consultée le 18/08/2020. Accessible au [https://www.google.com/intl/en\\_uk/search/howsearchworks/algorithms/](https://www.google.com/intl/en_uk/search/howsearchworks/algorithms/)

[5] PARISER Eli. The filter bubble - How the New Personalized Web Is Changing What We Read and How We Think. Penguin Books. 2012.

[6] Ciblage comportemental. Ryte.com. Accessible au : [https://fr.ryte.com/wiki/Ciblage\\_comportemental](https://fr.ryte.com/wiki/Ciblage_comportemental)

[7] ADEME, La face cachée du numérique, 2019

[8] ADEME, J. Lhotellier, E. Less, E. Bossanne, S. Pesnel. 2018. Modélisation et évaluation ACV de produits de consommation et biens d'équipement – Rapport. 186 pages.

## QU'EST-CE QU'UN

# Méta-moteur de recherche ?

Moteurs et méta-moteurs visent le même but : renvoyer un ensemble de résultats pertinent à une requête formulée par l'utilisateur. Ils se distinguent cependant au niveau de leur mode de fonctionnement. En effet, un **métamoteur est un logiciel qui puise ses informations à travers plusieurs moteurs de recherche** (notamment Google et Bing, mais également des moteurs spécialisés). Ainsi, le métamoteur explore les moteurs qui le sous-tendent en envoyant ses requêtes à ceux-ci, et retourne les résultats de chacun d'eux.

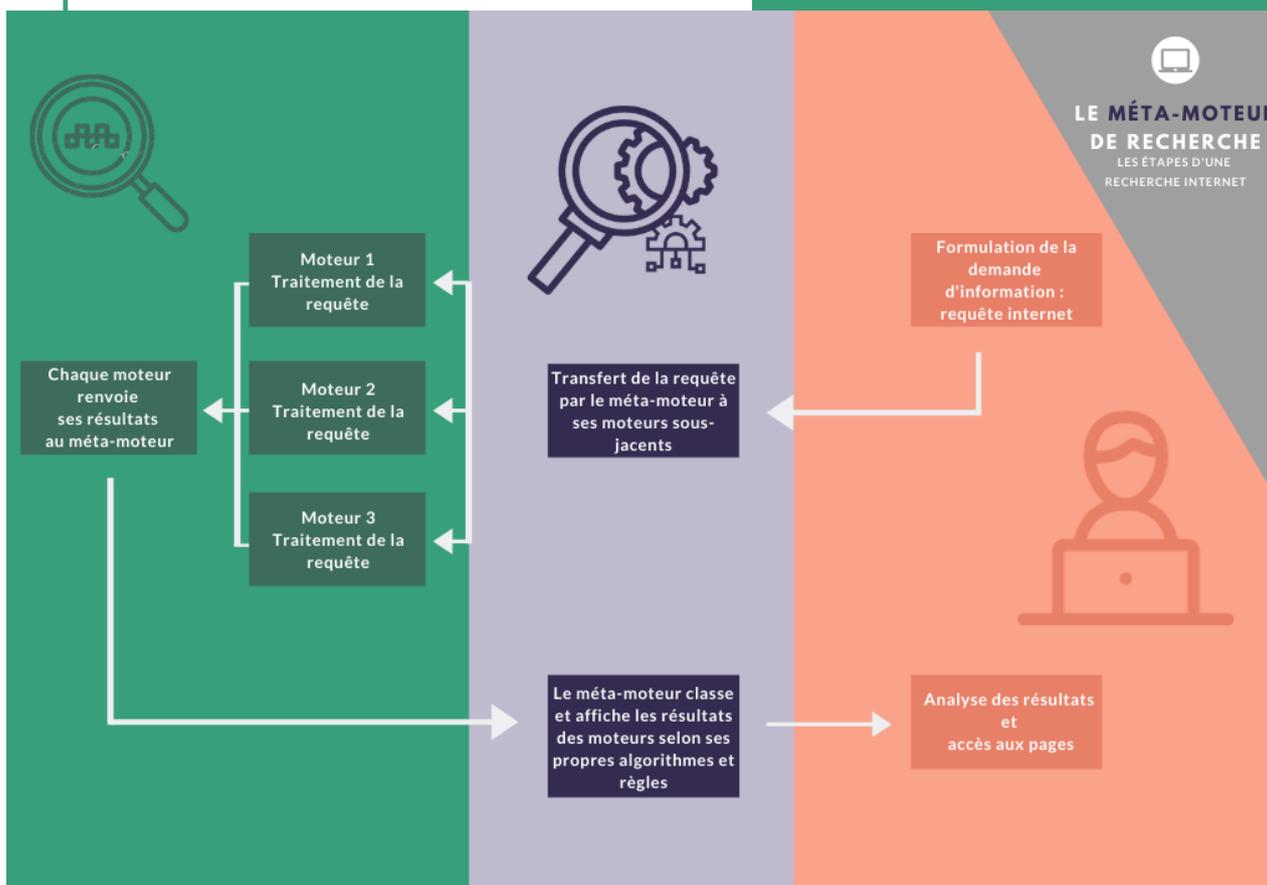
**Un des principaux avantages des méta-moteurs est de permettre aux utilisateurs de n'entrer leur recherche qu'une seule fois tout en accédant aux réponses de plusieurs moteurs de recherche différents** ; ces réponses étant traitées pour, d'une part, éliminer les résultats similaires ou liens en doublon et, d'autre part, remonter les résultats les plus pertinents pour chacun des moteurs de recherche[9]. Les métamoteurs agissent comme un intermédiaire entre les moteurs et l'internaute : les **informations personnelles** de celui-ci **ne sont pas communiquées aux moteurs, ce qui permet de remonter des résultats n'étant pas soumis au ciblage comportemental.**

Cependant, les métamoteurs présentent également quelques inconvénients.

1. A la différence des moteurs, ils ne possèdent pas leur index propre. Ils sont ainsi dépendants des informations fournies par des tiers (les moteurs) et ne peuvent obtenir que les données que ceux-ci acceptent de leur transmettre. **Les recherches peuvent ainsi être moins précises que sur un moteur performant.**
2. **Ils tendent à faire remonter principalement les premiers résultats trouvés sur les moteurs appelés.** Ceci peut créer des phénomènes pervers de prime aux sites les plus connus, d'auto-renforcement de la notoriété, de mise à l'écart de contenus plus confidentiels ou hétérodoxes...
3. **D'un point de vue environnemental,** la force d'un métamoteur devient une faiblesse. Au lieu de n'avoir recours « qu'à » l'infrastructure d'un moteur unique comme avec une simple recherche Google ou Bing, **l'utilisation d'un méta-moteur aura recours à l'infrastructure du méta-moteur en lui-même ainsi que celles des différents moteurs qu'il interroge[10].**

[9] Moteur de recherche, métamoteur et annuaire, quelles différences ? Culture-formations.fr, 2017. Accessible au : <https://culture-formations.fr/moteur-recherche-annuaire-metamoteur/>  
[10] Frédéric Bordage, Sobriété numérique, les clés pour agir, Buchet Chastel, 2019, 206p.

## LES ÉTAPES CONSTITUANT UNE RECHERCHE INTERNET



## LES GRANDS ACTEURS

### LES LEADERS MONDIAUX

	GOOGLE	BING	YAHOO
<b>CATÉGORIE</b>	 Moteur	 Moteur	 Moteur
<b>PARTS DE MARCHÉ <sup>[11]</sup> (MONDE, JUILLET 2020)</b>	92.2%	2.8%	1.6%
<b>ZONE D'INFLUENCE MAJORITAIRE</b>	Monde	Monde	Monde
<b>PROMESSE (ISSUE DU SITE DE L'ENTREPRISE)</b>	"Organiser les informations à l'échelle mondiale pour les rendre accessibles et utiles à tous"	"Fournir aux utilisateurs les résultats de recherche les plus pertinents à partir du Web, afin d'accéder facilement à du contenu de qualité produit par les éditeurs Web"	"Avec Yahoo, suivez les sujets qui vous passionnent, de l'actualité au divertissement, du sport à la finance"

### LES LEADERS LOCAUX

	BAIDU	YANDEX
<b>CATÉGORIE</b>	 Moteur	 Moteur
<b>PARTS DE MARCHÉ <sup>[12]</sup> (MONDE, JUILLET 2020)</b>	0.9%	0.9%
<b>ZONE D'INFLUENCE MAJORITAIRE</b>	Chine	Russie + pays voisins
<b>PROMESSE (ISSUE DU SITE DE L'ENTREPRISE)</b>	"Rendre la complexité du monde accessible à travers la technologie." [13]	"Notre but est de faciliter la navigation en ligne et hors ligne des consommateurs et professionnels." [14]

[11] [12] Données monde StatCounter.com, juillet 2020

[13] Version originale en anglais: "Making the complicated world simpler through technology"

[14] Version originale en anglais: "Our goal is to help consumers and businesses better navigate the online and offline world"

## LES GRANDS ACTEURS

### LES ALTERNATIVES SOLIDAIRES

	ECOSIA	LILO
<b>CATÉGORIE</b>	 Méta-moteur	 Méta-moteur
<b>ZONE D'INFLUENCE MAJORITAIRE</b>	Allemagne, USA, France, UK	France
<b>PROMESSE (ISSUE DU SITE DE L'ENTREPRISE)</b>	"Surfez sur le web tout en plantant des arbres"	"Vous permettre de financer gratuitement des projets solidaires & environnementaux grâce à vos recherches internet"
<b>PARTICULARITÉS</b>	Méta-moteur allemand. Consacre au moins 80 % des bénéfices issus de ses revenus publicitaires à la plantation d'arbres. Résultats provenant de Bing.	Méta-moteur français. Reverse 50% de ses revenus à des projets solidaires & environnementaux. Résultats provenant notamment de Google, Bing et Yahoo. Aucune collecte ou revente des données personnelles.

### LES ALTERNATIVES PROTÉGEANT NOS DONNÉES

	QWANT	DUCKDUCKGO	SEAR X
<b>CATÉGORIE</b>	 Moteur	 Méta-moteur	 Méta-moteur
<b>ZONE D'INFLUENCE MAJORITAIRE</b>	France, Allemagne	USA, UK, Allemagne, Canada	Monde
<b>PROMESSE (ISSUE DU SITE DE L'ENTREPRISE)</b>	"Protège les libertés de ses utilisateurs et veille à préserver l'écosystème numérique"	"Fatigués d'être traqués en ligne ? On peut vous aider"	"Un méta-moteur de recherche hackable et respectueux de la vie privée"
<b>PARTICULARITÉS</b>	Aucune collecte ou revente des données personnelles. Dispose de son propre moteur + complète ses résultats avec Bing. Moteur par défaut de l'administration publique française.	Aucune collecte ou revente des données personnelles. Résultats issus de nombreuses sources, notamment Bing, mais pas Google. Engagé dans l'open source.	Méta-moteur libre, décentralisé et paramétrable. Aucune collecte ou revente des données personnelles. Possibilité de choisir les moteurs sources (Bing, Google, Yahoo...).



Toutefois, de même que la meilleure énergie est celle que l'on ne consomme pas, **la recherche internet avec le moins d'impact reste celle que l'on ne fait pas**. Il existe ainsi **quelques conseils permettant de réduire facilement l'impact de ses recherches internet** :

- **Utiliser des favoris**
- **Taper directement l'adresse du site web recherché**
- **Utiliser des mots-clés les plus précis possibles**

Les deux premiers conseils permettent d'éviter une recherche internet lorsqu'elle n'est pas nécessaire tandis que le troisième optimise la recherche afin de moins solliciter les serveurs du/des moteurs de recherche.

Ainsi, il ressort de notre exploration que chacune de nos recherches internet, loin de sortir d'une forme de néant comme par magie une fois une requête formulée, se repose sur un vaste système incluant des processus informatiques complexes (crawl, index, classement...) ainsi que l'infrastructure physique supportant l'ensemble d'internet (terminaux, réseaux, datacenters). Nous avons ainsi pu ouvrir une fenêtre sur **la matérialité souvent oubliée du monde numérique**.

Dans ce domaine de la recherche internet Google émerge comme le leader incontesté, notamment du fait de la qualité de ses algorithmes et par conséquent des résultats affichés. Nous avons vu cependant qu'il est bon de garder quelques distances vis-à-vis des résultats affichés, des problématiques de **biais algorithmique, de ciblage comportemental ou encore de « bulle de filtre »** pouvant les influencer.

Google fait cependant face à de nombreux acteurs : des moteurs de recherche généralistes comme lui disposant de leurs propres algorithmes et index (Bing, Yahoo, Qwant...), et des méta-moteurs qui ajoutent une couche numérique à la recherche web, mais incluent également une considération solidaire et/ou de protection des données à leurs activités.

Ce rapport a été rédigé par la Communauté Do IT Green, qui a fêté ses deux ans d'existence au sein d'Informatique Caisse des Dépôts en 2020 et vise à rassembler les collaborateurs ICDC autour des sujets RSE.