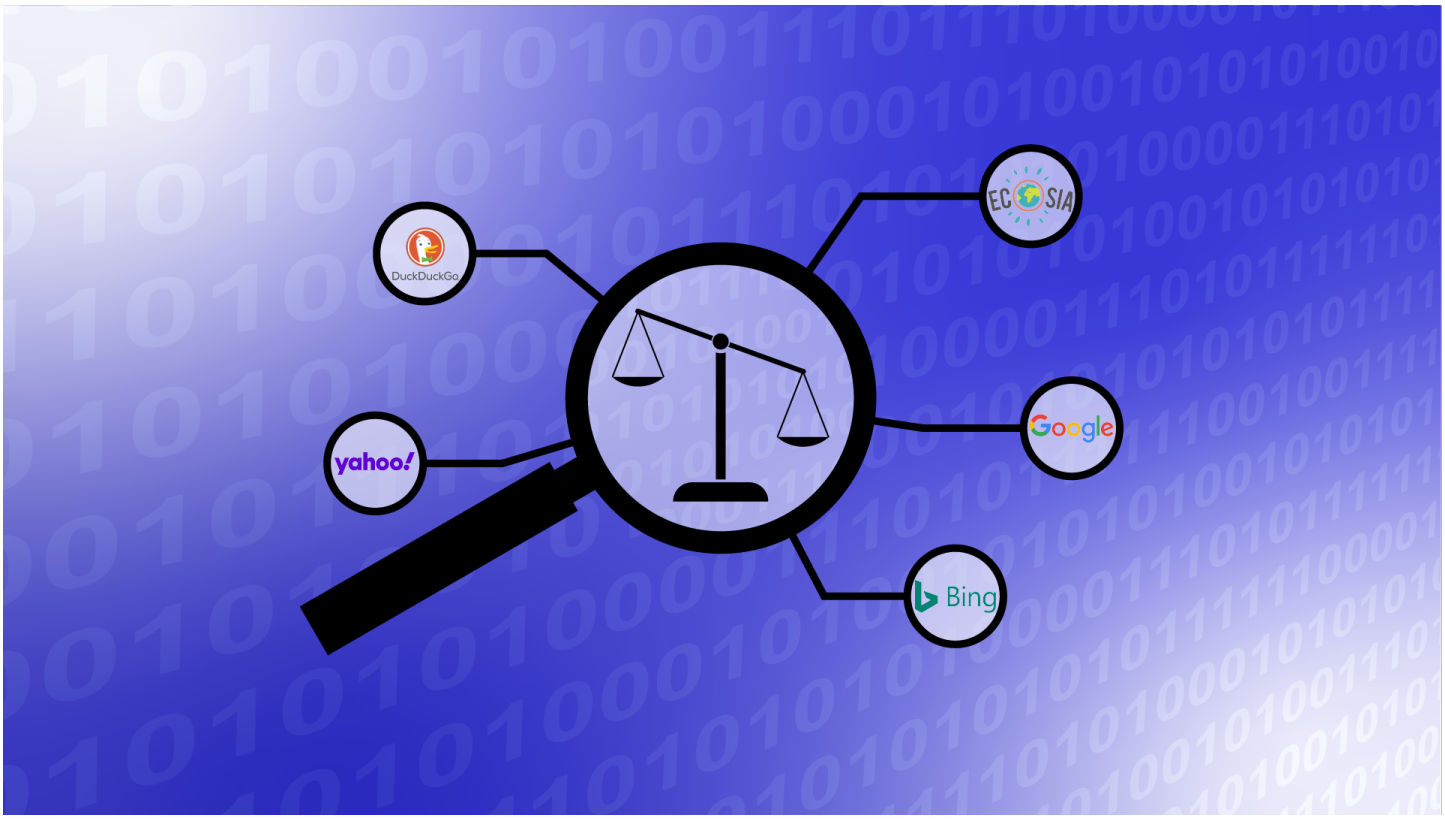
# Tester les biais des moteurs de recherche : pourquoi et comment ?

Guillermo Andrade Barroso, Patrick Maillé, Bruno Tuffin




**Publié le : 06/10/2021**

Par : Guillermo Andrade-Barroso, Patrick Maillé & Bruno Tuffin

**Niveau** ○○○

Niveau 2 : Intermédiaire



# Tester les biais des moteurs de recherche : pourquoi et comment ?

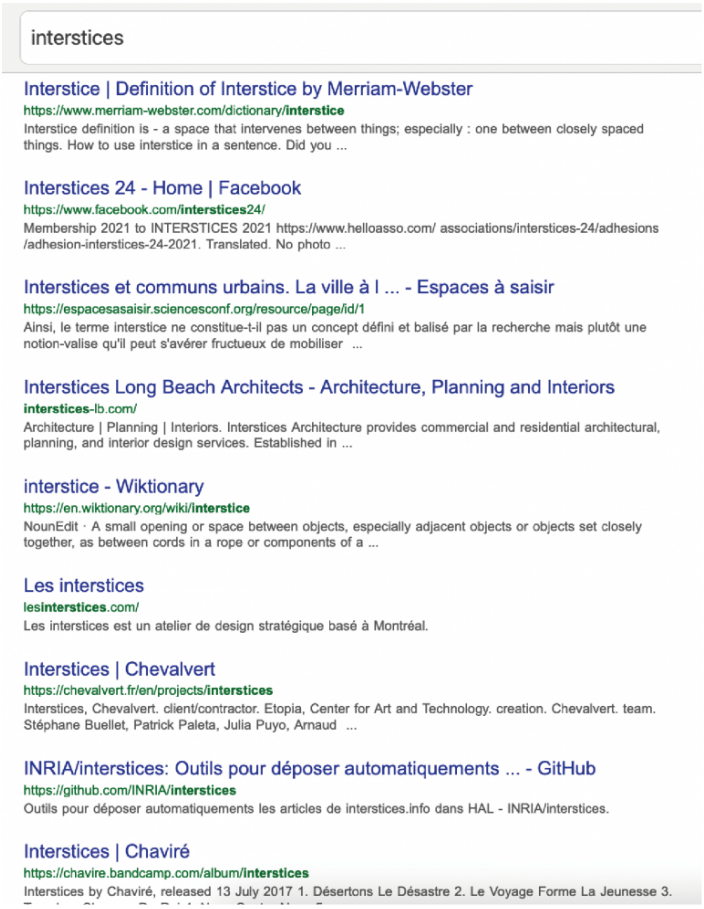
**CULTURE & SOCIÉTÉ**

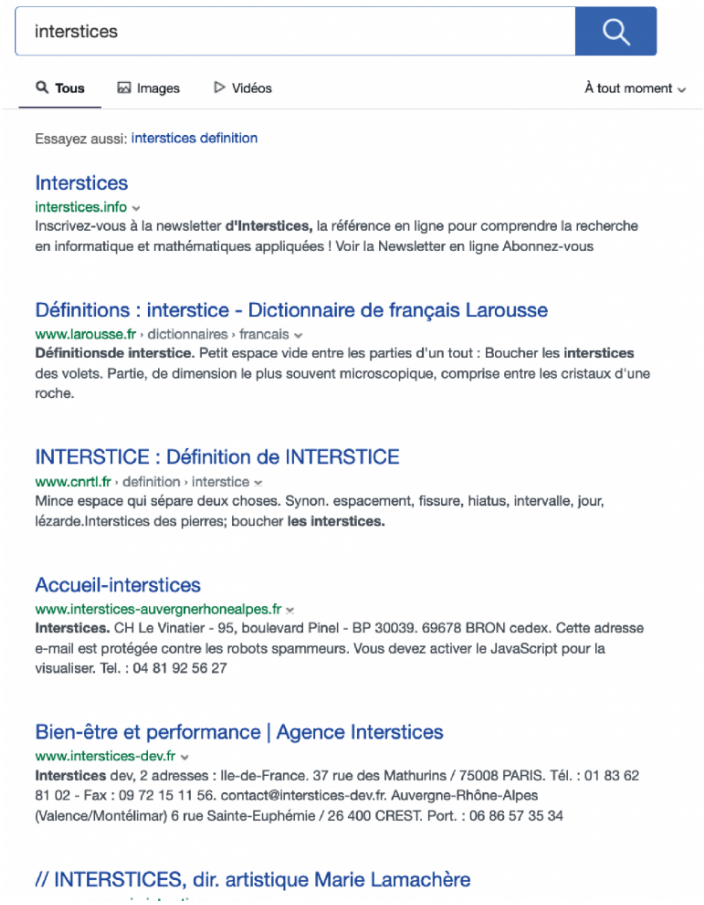
ALGORITHMES

*# Société* *# Recommandation* *# Google*

**Qwant, Ecosia, DuckDuckGo, Bing, Google... Nous avons tous un moteur de recherche préféré, que nous utilisons tous les jours pour accéder à des informations ou des services. Le principal critère à prendre en compte est son efficacité : sa capacité à fournir la bonne réponse à notre question. Mais ce moteur de recherche a ainsi une importante responsabilité : en choisissant une réponse plutôt qu'une autre, il définit notre expérience du Web ! Il est donc capital qu'il se base sur des critères objectifs et rationnels, et pas sur des intérêts privés ou des volontés idéologiques. S'assurer que les résultats d'un moteur de recherche ne sont pas biaisés, c'est le sujet de recherche de Guillermo Andrade-Barroso, Patrick Maillé et Bruno Tuffin, qui nous apportent quelques éléments de réponse...**

Les moteurs de recherche jouent un rôle central dans la manière dont on aborde le contenu sur Internet : quand on compose un mot clé sur son moteur favori, on reçoit en retour une liste de liens vers des pages ou contenus, liste supposée ordonnée selon des critères de pertinence. Les moteurs présentent une ergonomie différente, mais le premier lien a habituellement une probabilité plus forte d'être « cliqué » que le second, et ainsi de suite. Et non seulement la liste présentée a une importance, mais l'ordre donné également. La liste et son ordonnancement ont donc un impact majeur sur le contenu qui sera accédé.



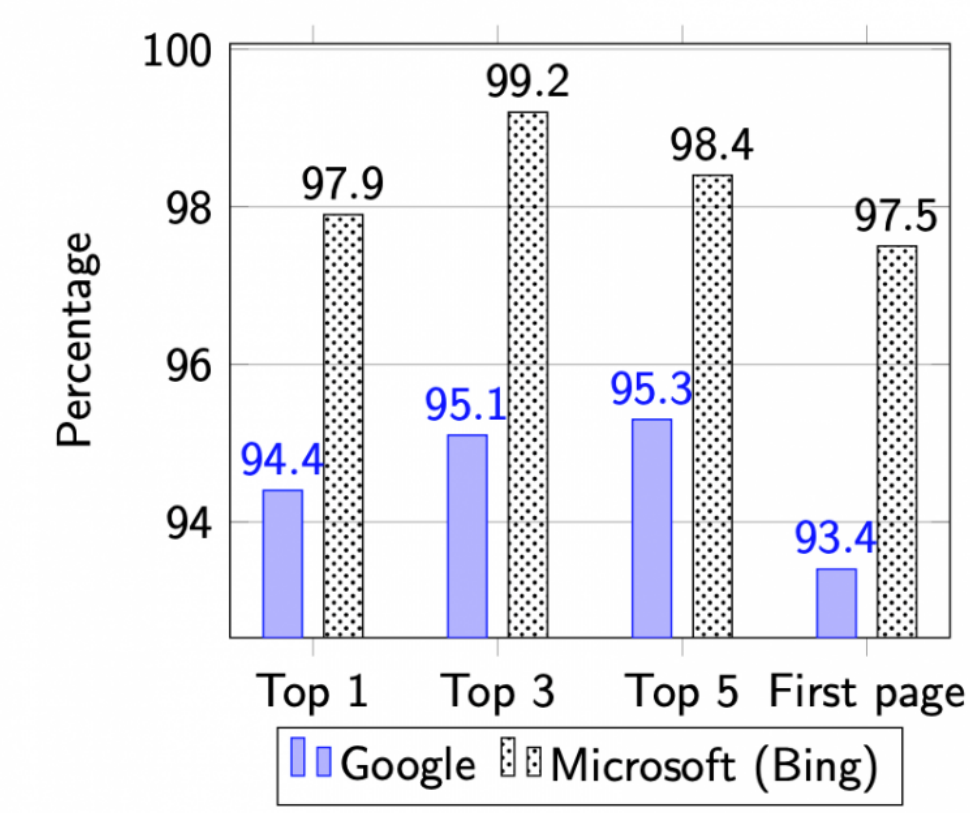

Plusieurs questions se posent alors, notamment :

Les moteurs ne donnent pas la même liste ordonnée pour un mot clé donné. Est-ce néfaste ? Est-ce au contraire utile ?

Les différences sont-elles volontaires pour favoriser certains contenus et orienter pour des raisons mercantiles (ou autres) vers certains sites ?

Pour la première question, une réponse semble *a priori* aller dans un sens positif : des choix différents peuvent avoir pour origine des notions de pertinence différentes et/ou une adaptation aux spécificités de l'utilisateur via la connaissance de son utilisation passée d'Internet (les **cookies**). La diversité semble donc être bénéfique pour permettre à l'usager de choisir le moteur de recherche le plus adapté, avec l'argument souvent invoqué par les moteurs que tout utilisateur n'est qu'à un clic de distance (*one click away*) d'un nouveau moteur s'il est mécontent des résultats retournés. Néanmoins, pour savoir si on préfère un moteur à un autre, il faudrait une analyse comparée approfondie des résultats obtenus, ce qui est raisonnablement réalisé par peu de personnes en pratique. De plus, le marché des moteurs est en fait peu compétitif, avec Google possédant plus de 92% du marché en avril 2021 (voir **ici** pour des chiffres complets mis à jour), et il est rare pour le commun des mortels de comparer et changer de moteur.

Ceci nous amène donc à répondre à la seconde question : Les moteurs de recherche biaisent-ils leurs résultats ? Et cela peut-il avoir une incidence sur l'utilisation du réseau Internet tout entier ? Cette question est devenue majeure quand notamment Adam Raff, le co-fondateur de la société de comparaison de prix Foundem, **s'est plaint en 2009** que son site était systématiquement mal classé par Google, qui favorisait ainsi ses propres services ; ceci pouvant être vu comme une atteinte à la concurrence. Des **études** semblent montrer que des sociétés comme Google ou Microsoft (dont le moteur est Bing) classent « leur » contenu bien mieux que les moteurs rivaux. C'est ce qu'on peut observer dans le diagramme suivant (source : **Wright 2012**), montrant le pourcentage de recherches via Bing ou Google dans lesquelles leur propre contenu classé chez eux n'est pas aussi bien classé par les concurrents.




Par exemple sur la troisième colonne, quand Bing classe du contenu Microsoft dans son top-5 pour une recherche, alors dans 98,4% des cas les concurrents ne le classent pas dans le top-5 pour la mê

valeurs étant dans le tableau toujours au-dessus de 90%, cela semble indiquer une forte tendance à favoriser son propre contenu.

Tout ceci a lancé un débat sur la neutralité des moteurs de recherche (*search neutrality* en anglais) dans la même veine que le débat sur la neutralité du Net. Pour la neutralité du Net, les fournisseurs d'accès sont ou étaient accusés de brider ou de différencier certains services, officiellement pour des raisons de congestion et fournir une qualité optimisée mais aussi pour inciter les fournisseurs de contenu à participer aux frais de maintenance du réseau. Ceci a soulevé de nombreuses inquiétudes sur l'avenir de l'Internet et l'impact sur l'innovation si le réseau n'était plus équitable et des contenus rendus plus difficilement accessibles. Depuis, en Europe notamment, la **Neutralité du Net** est protégée par la loi et surveillée par les régulateurs de télécommunication. Mais les problèmes d'accessibilité ne sont, comme on a pu le voir, pas seulement liés à la qualité d'accès des fournisseurs Internet : les moteurs de recherche jouent également un rôle qu'il ne faut pas minimiser et que certains estiment qu'il faudrait aussi réguler.

Ce débat sur la neutralité des moteurs est très sensible avec des réactions parfois très vives, plus encore que pour la neutralité du Net. Notre but est ici non pas d'influencer le lecteur dans un sens ou un autre, chacun étant libre de son opinion, mais de donner des éléments et méthodes de contrôle de biais éventuels.

## Est-il possible de mettre en évidence des biais ?

Alors qu'établir qu'un fournisseur d'accès Internet différencie les services n'est pas toujours simple, prouver ou étudier les biais d'un moteur pour une recherche donnée est probablement encore plus complexe. En effet, une telle étude requiert l'utilisation d'une fonction objective de qualité qui n'est pas aussi simple à définir pour les recherches : comment définir la pertinence d'un résultat, et pourquoi y aurait-il une seule notion possible et adéquate ?

Ce problème peut paraître insoluble. L'option que nous allons suivre est de récolter les résultats de plusieurs moteurs de recherche pour étudier si un ou des résultats sur un moteur semblent aberrants par rapport à ce qui est affiché par les autres ; ceci pourrait être une indication de biais volontaire potentiel, mais en aucun cas une preuve, juste un indicateur suggérant que des investigations supplémentaires peuvent être réalisées. C'est le but de l'action exploratoire Inria **SNIDE**.

## Méthodologie et tests statistiques

Plutôt que de définir une notion (subjective) de pertinence, nous allons utiliser comme mesure la *visibilité* que les moteurs de recherche accordent aux liens. Cette visibilité peut être supposée proportionnelle à la probabilité de clic associée à chaque position, qui correspond à l'intérêt (relatif) éveillé par cette même position ; on considère donc sans perte de généralité comme visibilité cette probabilité de clic. Ces probabilités peuvent être estimées en comptant, sur un grand nombre de recherches, le nombre de clics pour chaque position

Utiliser une telle mesure est algorithmiquement simple et prend en compte le niveau d'accessibilité d'un contenu : on quantifie alors un changement dans un classement en termes de visibilité via un changement de probabilité d'être accédé par un clic. Les probabilités utilisées pour les calculs peuvent être choisies arbitrairement, mais nous utiliserons les valeurs pour les 10 premières positions sur un moteur de recherche (0.364, 0.125, 0.095, 0.079, 0.061, 0.041, 0.038, 0.035, 0.03, 0.022) obtenues selon des **tests intensifs**.

La visibilité d'une page sur un moteur sera de valeur zéro si la page n'est pas affichée, et la visibilité globale d'une page sera définie comme la moyenne des visibilités pour cette page sur tous les moteurs de recherche. On peut alors également donner un *score* à un moteur de recherche pour un mot clé en sommant les visibilités des pages affichées, mais en pondérant ces visibilités par leurs probabilités de clic associées à leurs positions sur ce moteur : on quantifie ainsi le fait qu'un moteur de recherche a tendance à montrer des pages « populaires » (au sens qu'elles sont montrées par l'ensemble des moteurs), i.e., qu'il est plutôt en accord avec les autres moteurs.

Tout ce que nous décrivons est implanté sur le site **https://snide.irisa.fr/**. Quinze moteurs de recherche sont comptabilisés, listés ci-dessous, avec leurs parts de marché en mai 2021 selon le site **https://gs.statcounter.com** :

| Nom | Logo | Part de marché France | Part de marché Monde | Commentaire |
|---|---|---|---|---|
| AllTheInternet | All the Internet<br>A better way to search the Internet. | <0.01% | <0.02% | |
| AOL | Aol. | <0.01% | <0.02% | Basé sur Bing depuis 2015 |
| Ask | Ask | <0.01% | <0.02% | |
| Bing | | 3.77% | 2.27% | Le moteur de recherche de Microsoft |
| DirectHit | DIRECT HIT | <0.01% | <0.02% | |
| DuckDuckGo | | 0.51% | 0.59% | Vise à préserver la vie privée et éviter les bulles de filtre |
| Ecosia | ECOSIA | 1.11% | 0.13% | Basé sur Bing |
| Google | | 91.99% | 92.2% | En position dominante en Fran |

| | | | | |
|---|---|---|---|---|
| | | | | monde |
| Latlas | | <0.01% | <0.02% | Hébergé en France, n'utilise ni historique, ni localisation géographique, sans publicité |
| Lilo | | <0.01% | <0.02% | Méta-moteur basé sur Bing et Google |
| Lycos | | <0.01% | <0.02% | |
| Qwant | | 0.98% | 0.03% | Objectif de protection de la vie privée |
| Startpage | | <0.01% | <0.02% | |
| Teoma | | <0.01% | <0.02% | |
| Yahoo | | 1.36% | 1.5% | Basé sur Bing |
| Yandex | | <0.01% | 1.21% | Basé en Russie, environ 50% du marché en Russie |

Découvrez ci-dessous l'application de tels calculs et des scores associés dans l'interface de **SNIDE**, pour l'exemple d'une recherche du mot « algorithme ».


> *Lorsqu'on effectue une recherche sur l'outil, par exemple pour le mot-clé « algorithme », celui-ci récupère les dix premiers liens organiques retournés par chaque moteur de recherche, qu'on peut visualiser d'un simple clic. Chaque page accumule de la visibilité selon sa position chez un moteur via les valeurs listées ici : la première place rapporte 0.36 points de visibilité, la seconde 0.125, et ainsi de suite. La moyenne obtenue par chaque page sur l'ensemble des moteurs de recherche est affichée à côté du lien. Si on classe les pages selon cette visibilité moyenne, on obtient le classement appelé « consensus ».*

Nous pouvons associer une notion de visibilité aux pages affichées selon les moteurs. Mais comment déterminer si certaines pages sont étrangement plus ou moins affichées par un moteur ? Des techniques de détection de valeurs aberrantes (*outlier detection* en anglais) peuvent être utilisées. Nous proposons ici d'utiliser des outils statistiques basés sur le **test de Dixon**, un test d'hypothèse standard en statistiques. Ce test identifie des valeurs aberrantes parmi un échantillon de petite taille : il détermine si une valeur extrême (la plus grande ou la plus petite) est aberrante par rapport aux autres. Nous l'appliquons pour tester quatre hypothèses différentes :

Pour un mot clé, si le score d'un moteur est anormalement faible, et donc si le moteur affiche des résultats très différents des autres.

Si la page la plus visible pour le mot clé est classée anormalement bas par un moteur, qui pourrait signifier que le moteur déclasse cette page.

Si un moteur donne abusivement la première place à une page non considérée pertinente par les autres moteurs.

Pour chaque moteur, si la page classée première est également bien classée par les concur

L'application rigoureuse du test de Dixon suppose des hypothèses sur les valeurs utilisées pour l'échantillon, ici le score d'un moteur ou la probabilité de clic des positions selon l'hypothèse testée (valeurs supposées indépendantes et distribuées selon une loi gaussienne). Ces hypothèses probabilistes ne sont certainement pas vérifiées, mais les tests permettent néanmoins de pointer des valeurs à étudier de plus près.


*L'outil calcule un score pour chaque moteur, qui est simplement la somme des scores des pages montrées, pondérés par la visibilité de chaque position. Pour le mot-clé « algorithme », les scores des moteurs de recherche sont assez proches les uns des autres, et aucun n'est d'ailleurs relevé par le test statistique correspondant, dont les résultats sont indiqués en haut à droite.*
*Par contre, la page montrée en première position par plusieurs moteurs a un score assez faible, en ce sens qu'elle n'est pas beaucoup montrée par les autres. C'est ce qui est relevé par le dernier test statistique. Par exemple on remarque dans notre classement consensus que la page la plus visible en moyenne est la page Wikipedia pour « Algorithme », alors que AOL et Yahoo nous envoient vers la page Wikipedia anglophone. On peut difficilement les accuser ici d'avoir un comportement malhonnête, au plus remarquer qu'ils pousseront peut-être davantage des contenus anglophones.*
*Les autres moteurs qui mettent en tête de leur classement une page peu montrée par les autres sont Ask, qui montre une vidéo YouTube, ainsi que cinq autres moteurs. On peut alors regarder le détail pour décider si ces cas mis en évidence par l'outil peuvent venir d'un classement qui ne serait pas basé sur la seule pertinence des pages pour l'utilisateur.*

# Méta-moteurs

En dehors de tests statistiques, recenser les résultats de différents moteurs et associer une visibilité aux pages permet également de construire des méta-moteurs (c'est-à-dire des moteurs agrégeant les résultats) qui permettent sinon d'éliminer les biais, tout du moins de les lisser.

Un premier méta-moteur que nous pouvons appeler le *moteur consensus*, retourne pour un mot clé les pages ordonnées selon leur visibilité moyenne (qui rappelons-le est la moyenne des visibilités obtenues sur les différents moteurs) : la page de plus haute visibilité d'abord, et ainsi de suite.

Un second moteur, appelé *moteur jugement majoritaire*, utilise la **méthode utilisée pour les votes** et, schématiquement, ordonnant les pages non selon leur visibilité moyenne mais selon leur visibilité médiane (c'est-à-dire la valeur de visibilité située au milieu de toutes les valeurs obtenues par la page aux différents moteurs de recherche). Ce principe, ainsi que différentes méthodes de vote existantes, ont déjà été décrit dans un article paru sur Interstices **ici**. En procédant ainsi, un biais volontaire augmentant ou diminuant la visibilité sur un moteur ne provoquera au pire qu'un décalage d'un cran pour la valeur milieu parmi toutes les visibilités, voire n'aura aucune incidence. Cette technique devrait permettre de lisser les biais plus qu'en utilisant le moteur consensus. Ceci a été **observé sur des données synthétiques**, et est également illustré ci-dessous.


*Dans le classement consensus, on retrouve des pages qui ne sont montrées que par un seul moteur de recherche, si elles sont placées assez haut pour avoir un score de visibilité qui reste c*

*On a le cas ici pour trois pages, qui sont chacune en première position chez un moteur mais ne sont montrées par aucun autre. Le classement consensus reste donc fragile si des moteurs voulaient artificiellement mettre une page en avant. À l'inverse, le classement basé sur le principe du jugement majoritaire ne montre que les pages qui sont montrées par plusieurs moteurs, car le classement raisonne sur la visibilité médiane (et non moyenne) des pages. On s'attend donc à ce qu'il soit plus robuste que le classement consensus vis-à-vis de biais qu'un moteur de recherche pourrait avoir. Les trois pages mentionnées n'apparaissent d'ailleurs pas dans ce nouveau classement.*

# Conclusion

Le débat sur la neutralité des moteurs de recherche et l'influence des résultats est très vif et sensible. Sans prendre position, il est nécessaire d'avoir des outils à disposition permettant le plus simplement possible de pointer des résultats suspicieux, mais qui ne sont pas nécessairement synonymes de biais volontaires, les moteurs pouvant utiliser des algorithmes avec des objectifs et des paramètres différents. La méthode simple présentée ici utilise les visibilités des différentes pages sur les moteurs et permet d'identifier via des tests statistiques si un résultat paraît suspect. La réduction des biais peut être obtenue via les méta-moteurs que nous avons présentés et implantés sur le site **snide.irisa.fr**.