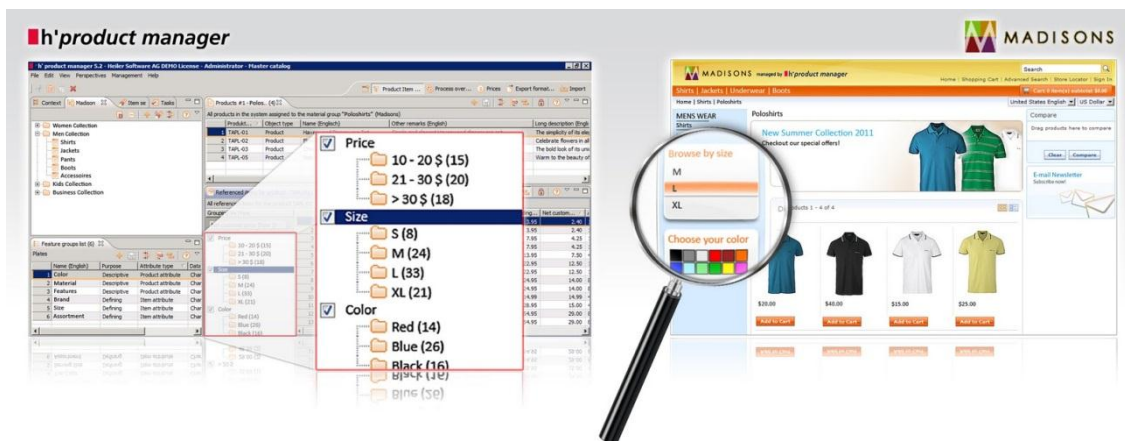


LIVRE BLANC

SEO pour les sites marchands

Optimiser une navigation à facettes

01 septembre 2013
Version 1.1



Search Foresight
Groupe My Media
55 rue Anatole France 92300 LEVALLOIS PERRET

Contact : Philippe YONNET
Tél. : +33 1 74 18 29 40
Mob : +33 6 99 60 21 49
philippe.yonnet@search-foresight.com

Table des matières

A.	Préambule :	4
A.1	A propos de Search Foresight	4
A.1.1	Search Foresight, le spécialiste de l'accompagnement stratégique en SEO	4
A.1.2	Pourquoi « Search Foresight » ?	5
A.1.3	Une agence fortement intégrée à My Media	6
A.2	Pourquoi ce livre blanc ?	7
B.	Qu'est-ce qu'une facette ? Et qu'est-ce qu'une navigation à facettes ?	8
B.1	Les « facettes » sont des filtres de recherche composés de combinaisons valeurs x attributs	8
B.2	De la recherche à facettes à la navigation à facettes	9
C.	Comment une navigation à facettes impacte le SEO	10
C.1	Facettes et « pièges à robots » (« spider traps »)	10
C.1.1	un crawl perturbe par l'explosion combinatoire du nombre de pages de listing	11
C.1.2	la plupart des pages ainsi filtrées ne renvoient pas de résultats	12
C.1.3	Navigation à facettes et contenu dupliqué	13
C.1.4	les facettes vues par les moteurs	13
C.1.5	Matt cutts et le contenu à facettes	16
D.	Optimiser sa navigation à facettes	18
D.1.1	trier le bon grain de l'ivraie : identifier les « bonnes » facettes pour le seo	18
E.	Comment ouvrir ou fermer une « facette »	19
E.1	Gérer les différents états pour une facette	19
E.1.1	MISE A DISPOSITION DU FILTRE DANS L'interface utilisateur	19
E.1.2	ouverture des pages filtrées pour les moteurs de recherche (regles a priori)	19
E.1.3	REGLES De gestion GEREES A POSTERIORI pour les urls filtrées par facettes ouvertes au SEO	20
F.	Navigation à facettes et performances	22
F.1.1	optimiser les temps de téléchargement des pages filtrées par facettes	23
G.	Exploiter les facettes dans un moteur de recherche interne	23
G.1	Comment cela fonctionne-t-il ?	23
G.2	Les avantages	24
H.	Quels gains espérer de ces optimisations	24
H.1	Gain n°1 : amélioration de l'expérience utilisateur	24
H.2	Gain n°2 : plus de trafic capté	25



H.3	Gain n°3 : plus de conversions	25
I.	Ce que Search Foresight peut faire pour vous aider	26
J.	Conclusion : structurez toutes vos informations pour mieux les exploiter	27

A. Préambule :

A.1 *A propos de Search Foresight*

A.1.1 SEARCH FORESIGHT, LE SPECIALISTE DE L'ACCOMPAGNEMENT STRATEGIQUE EN SEO

Fondée par deux consultants sénior, l'agence Search Foresight met au service de ses clients une grande expérience :


- Du référencement de sites avec des enjeux SEO forts
- De problématiques techniques complexes et des fortes volumétries de données
- Du référencement de sites de marques
- Du référencement de sites internationaux, et/ou multilingues
- D'une clientèle exigeante, orientée vers les résultats



Le management de Search Foresight

A.1.2 POURQUOI « SEARCH FORESIGHT » ?

Nous avons choisi ce nom car il symbolise notre vision du SEO et de l'apport que nous voulons proposer à nos clients :

- **Prévision:** les algorithmes des moteurs de recherche évoluent de plus en plus rapidement, et les technologies également. La mission de notre agence est de proposer à nos clients des stratégies durables, qui ne seront pas obsolètes avant d'avoir été totalement mises en œuvre.
- 
- **Prévoyance :** nous nous devons de ne proposer que des approches parfaitement sûres à nos clients
 - **Vision à long terme :** mettre à la disposition de nos clients dès maintenant les méthodes et les technologies qui seront employées par tous demain fait partie de notre ADN. C'est vrai en particulier des apports pour le référencement des sites de technologies avancées comme :
 - Le NLP : Natural Language Processing
 - La reconnaissance des entités nommées
 - Les technologies NoSQL et le BigData
 - Le web mining et le data mining

A.1.3 UNE AGENCE FORTEMENT INTÉGRÉE À MY MEDIA

Search Foresight, si elle est une agence autonome et fortement spécialisée, fonctionne de manière très intégrée à l'agence My Media. C'est particulièrement vrai en accompagnement stratégique où les ressources et les compétences de My Media sont mises en œuvre pour déterminer la meilleure approche pour développer le trafic du client, tous canaux confondus.

Les équipes SEM et SEO travaillent en commun au quotidien. Cela permet notamment de pouvoir proposer des arbitrages pertinents sur les budgets SEM, notamment quand les progrès en SEO peuvent justifier une diminution des dépenses.



A.2 *Pourquoi ce livre blanc ?*

Chez Search Foresight, notre conception du conseil passe par [le partage du savoir](#). Nous pensons que plus nos clients [montent en compétence](#), plus ils nous confieront de missions intéressantes et avancées, qui sont celles qui nous passionnent. En plus nous sommes persuadés que nos chances d'avoir des résultats augmentent avec la capacité de nos clients à comprendre les techniques et les chantiers que nous proposons.

C'est la raison pour laquelle nous avons créé le [Centre de Formation Search Foresight](#), et que nous organisons régulièrement des [petits déjeuners d'information](#) sur les techniques SEO de pointe. Cela passe aussi par des conférences dans les événements e-marketing, de publication d'articles et ... de livres blancs comme celui-ci.

Nous avons choisi de rédiger un premier livre blanc sur la thématique de la navigation à facettes, car nous rencontrons trop souvent de mauvaises implémentations de cette technique.

Nous espérons que ce travail permettra de sensibiliser nos lecteurs sur les avantages et inconvénients de l'exploitation des facettes pour la navigation, et que nous pourrons convaincre certains d'entre vous du potentiel inexploité qui peut exister sur leurs sites...

B. Qu'est-ce qu'une facette ? Et qu'est-ce qu'une navigation à facettes ?

B.1 Les « facettes » sont des filtres de recherche composés de combinaisons valeurs x attributs

Le terme « **facette** » est apparu dans le jargon des moteurs de recherche pour désigner des requêtes filtrées sur des attributs de la classe d'items stockée dans l'index d'un moteur de recherche.

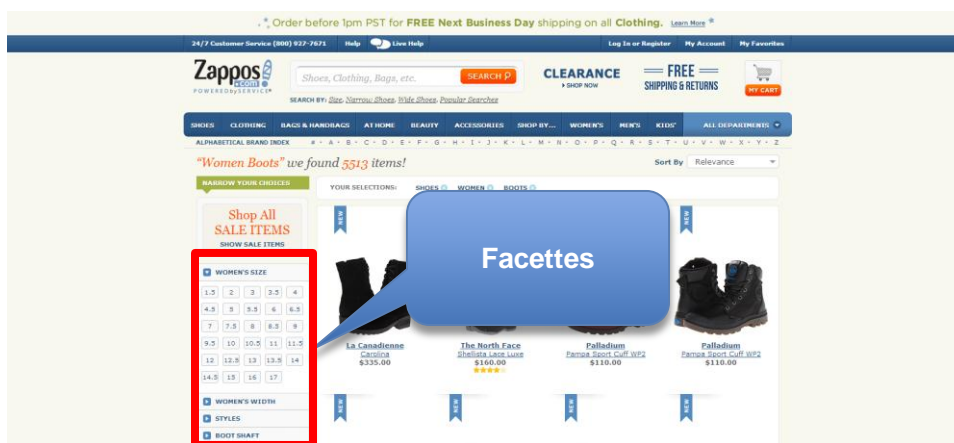
Dans le contexte d'un site marchand, la classe d'items évoquée plus haut est constituée par le catalogue des produits, et les attributs sont ses caractéristiques (couleur, matière, taille, poids, prix, puissance etc.)

Une facette est donc un filtre, associé soit à un couple {attribut,valeur} (ex couleur='rouge'), soit à plusieurs valeurs d'un même attribut (couleur='rouge' OU couleur='rose'). Mais on peut définir aussi les valeurs comme des bornes MIN, MAX, ou combinées (prix < 500 euros, ou prix compris entre 100 et 500 euros).

Les filtres attribut x valeur(s) ou attribut x borne(s) de valeur peuvent être combinées pour créer des **filtres composites** ([couleur='rouge' OU couleur='beige'] ET taille='40' ET matière='coton')

Le terme « facette » (*facet* en anglais) est né d'une analogie entre ce mode de recherche filtrée multidimensionnelle et les multiples « facettes » d'un diamant.

Un moteur de recherche optimisé pour répondre facilement sur une requête filtrée par attributs est donc prêt à examiner des collections de produits... sous toutes leurs facettes.



Un exemple de site utilisant les « facettes » de manière intensive : zappos.com

B.2 De la recherche à facettes à la navigation à facettes

Dès lors qu'ils disposent d'un moteur de recherche interne capable de gérer des filtres par attributs, les webmasters mettent logiquement en place une « navigation par facettes ».

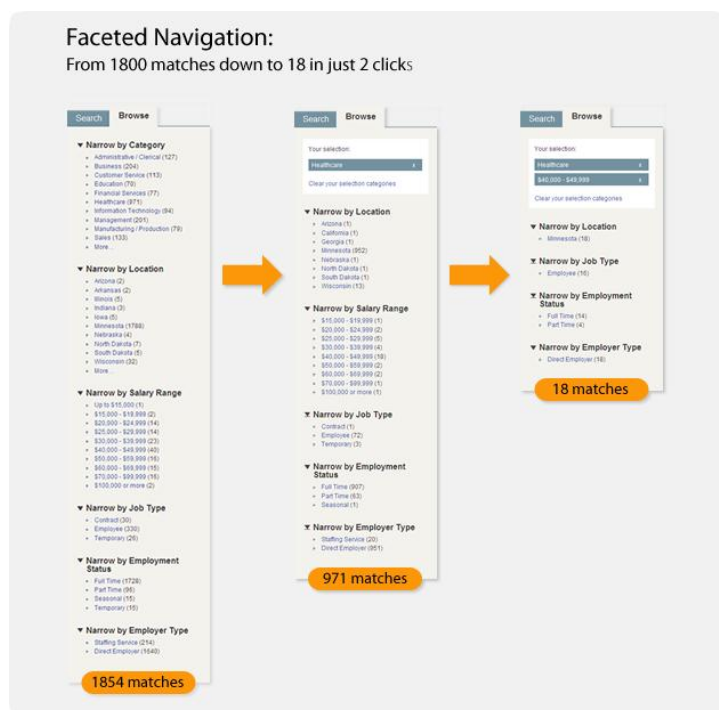
Dans la pratique, on crée une première navigation descendante dans le catalogue en suivant peu ou prou l'arborescence des catégories :

Home > Univers Produits > Catégorie > Sous-Catégorie >

Au niveau des catégories de plus bas niveau, il peut rester un grand nombre d'items (de modèles, de SKU, de produits) à afficher. Dans ce cas, on peut proposer à l'utilisateur de « filtrer » sa recherche en sélectionnant des valeurs de filtres proposés classiquement soit dans la colonne de gauche (en face des items), soit au-dessus des items.

La technique utilisée pour filtrer les résultats peut différer grandement d'un site à l'autre :

- Utilisation de **formulaires HTML**
- Utilisation de **liens**
- Utilisation de **scripts javascript**, avec éventuellement changement d'état de la page **en ajax** pour éviter de « recharger » toute la page...



Un exemple de l'expérience de navigation que crée une navigation à facettes : en deux clics, on peut restreindre le nombre des items qui correspondent aux critères de recherche d'un facteur 100 : de 1800 à 18 !

C. Comment une navigation à facettes impacte le SEO

Trop souvent, une mauvaise implémentation d'une navigation à facettes a des conséquences néfastes sur le SEO.

C.1 Facettes et « pièges à robots » (« spider traps »)

Si l'on vend des vêtements sur un site marchand, on s'aperçoit assez vite que pour des requêtes de type « pull rouge », la meilleure **page d'atterrissage** est une page de catégorie filtrée par valeur d'attribut : à savoir la catégorie « pulls » filtrée par « couleur='rouge' ».

On peut donc être tenté d'associer chaque page de catégorie filtrée par attributs avec une url.

Cette association : page facetée <-> url peut aussi se produire involontairement pour des raisons techniques, notamment si sur le site, les filtres sont appelés sous formes de paramètres passés en mode GET (c'est-à-dire : dans la chaîne de paramètres ou query string)

Pour Googlebot, le robot d'exploration de Google, cela signifie qu'il va trouver sur le site un nombre très grand d'urls avec des syntaxes de type :

www.lesitemarchand.com/univers/categ/sous-categ/?couleur='rouge'&filtreprix='<500'

ou

www.lesitemarchand.com/univers/?categ=1&couleur='rouge'&filtreprix='<500'

ou

www.lesitemarchand.com/univers/categ/sous-categ/couleur/rouge/filtreprix/moins500

ou ...

Selon la syntaxe utilisée sur le site...

Le problème ici, c'est que le nombre de combinaison entre facettes peut très vite faire exploser le nombre de pages...

Prenons un site comportant 2000 produits répartis dans 100 catégories... Classiquement, ce site comportera au départ au maximum 2500 pages, dont 2000 pages produits.

Maintenant, imaginons que ces produits soient des vêtements, disponibles dans 10 tailles, 5 couleurs, 5 matières, et dans 5 segments de prix, et provenant d'une trentaine de marques...

Chaque page de catégorie génère à présent potentiellement :

$1 \times 10 \times 5 \times 5 \times 5 \times 30 = 37500$ pages filtrées différentes ! Soit 3 750 000 pages de listings pour tout le site !

Et encore, le nombre d'urls peut être encore plus grand si on « permute » l'ordre des paramètres dans les urls, ce qui arrive souvent lorsque les facettes sont ouvertes par accident en passant les paramètres en mode GET...

C.1.1 UN CRAWL PERTURBE PAR L'EXPLOSION COMBINATOIRE DU NOMBRE DE PAGES DE LISTING

Le robot (Googlebot pour Google) va donc dépenser une énergie considérable à « crawler » (télécharger pour analyse) des pages de listing filtrées. Et si votre site n'est pas jugé assez « important » (ce qui se traduit par : n'a pas assez de Pagerank, et oui, ce critère est important pour le crawl), cela signifie probablement que certaines pages utiles de votre site ne vont pas être crawlées « normalement » (notamment certaines pages produits, et certaines des meilleures pages filtrées par attributs).

Ensuite il est évident que les millions de page ainsi créées ne sont pas toutes intéressantes à indexer pour les moteurs. En fait, l'ensemble des pages filtrées par facettes d'un site marchand ressemble... à du spam.

Objection : Ok, ok, j'ai une url par page filtrée par facette, mais ces urls ne sont pas présentes sur le site ! Donc Google ne peut pas les trouver...

Et si ! Dans la pratique, il suffit que Google découvre le lien via une page web externe (le blog d'un de vos clients qui recommande une page) ou via la Google Toolbar, et il crawlera la page... Cela va même plus loin : si la syntaxe de l'url est « comprise » par Google, son crawl peut s'étendre à d'autres valeurs de paramètres.

Dans le cas où les urls des pages de catégories filtrées par facettes ne sont pas présentes sur le site, le nombre de ces urls crawlées par Google sera beaucoup plus limité (mais très variable d'un cas à l'autre).

From: Webmaster Tools [mailto:wmt-noreply@google.com]
Sent: 07 December 2010 13:51
To: [REDACTED]
Subject: Googlebot found an extremely high number of URLs on your site:
[http://www.\[REDACTED\].co.uk/](http://www.[REDACTED].co.uk/)

Googlebot encountered problems while crawling your site [http://www.\[REDACTED\].co.uk/](http://www.[REDACTED].co.uk/).

Googlebot encountered extremely large numbers of links on your site. This may indicate a problem with your site's URL structure. Googlebot may unnecessarily be crawling a large number of distinct URLs that point to identical or similar content, or crawling parts of your site that are not intended to be crawled by Googlebot. As a result Googlebot may consume much more bandwidth than necessary, or may be unable to completely index all of the content on your site.

[More information about this issue](#)

Here's a list of sample URLs with potential problems. However, this list may not include all problematic URLs on your site.

- [http://www.\[REDACTED\].co.uk/co/pink,royal-blu,taupe,mustard,gold,khaki/](http://www.[REDACTED].co.uk/co/pink,royal-blu,taupe,mustard,gold,khaki/)
- [http://www.\[REDACTED\].co.uk/co/royal-blu,ch-brown,chestnut,fuscia/](http://www.[REDACTED].co.uk/co/royal-blu,ch-brown,chestnut,fuscia/)
- [http://www.\[REDACTED\].co.uk/co/turquoise,peach,charcoal,lt-denim,red/si:5/](http://www.[REDACTED].co.uk/co/turquoise,peach,charcoal,lt-denim,red/si:5/)
- [http://www.\[REDACTED\].co.uk/co/assorted,indigo,cerise,chestnut,charcoal/si:12/](http://www.[REDACTED].co.uk/co/assorted,indigo,cerise,chestnut,charcoal/si:12/)
- [http://www.\[REDACTED\].co.uk/co/peach,ch-brown,turquoise,coral,fuscia,khaki/](http://www.[REDACTED].co.uk/co/peach,ch-brown,turquoise,coral,fuscia,khaki/)
- [http://www.\[REDACTED\].co.uk/co/aqua,turquoise,assorted,indigo,chestnut,beige/](http://www.[REDACTED].co.uk/co/aqua,turquoise,assorted,indigo,chestnut,beige/)
- [http://www.\[REDACTED\].co.uk/co/silver,tan,taupe,peach/si:4/](http://www.[REDACTED].co.uk/co/silver,tan,taupe,peach/si:4/)
- [http://www.\[REDACTED\].co.uk/co/taupe,cerise,ch-brow,pink,blue/](http://www.[REDACTED].co.uk/co/taupe,cerise,ch-brow,pink,blue/)
- [http://www.\[REDACTED\].co.uk/clearance/p:0-20/sim:7/co:grey/](http://www.[REDACTED].co.uk/clearance/p:0-20/sim:7/co:grey/)
- [http://www.\[REDACTED\].co.uk/co/gold,assorted,brown,beige/sty:tunic/](http://www.[REDACTED].co.uk/co/gold,assorted,brown,beige/sty:tunic/)
- [http://www.\[REDACTED\].co.uk/co/red,white,grey,lt-denim,d-blue/](http://www.[REDACTED].co.uk/co/red,white,grey,lt-denim,d-blue/)
- [http://www.\[REDACTED\].co.uk/co/turquoise,gold,charcoal,teal,cream/no-](http://www.[REDACTED].co.uk/co/turquoise,gold,charcoal,teal,cream/no-)

Un exemple de message envoyé par Google via l'interface de Google Webmaster Tools signalant que Googlebot a rencontré des problème en recontrant « un nombre très grand de liens sur votre site ». Dans ce cas, le « piège à robots » est clairement généré par le nombre élevé d'urls créées par une navigation à facettes entièrement crawlable.

Notons qu'il y'a une méthode, décrite plus loin, pour « empêcher » Google de crawler et d'indexer ces urls.

C.1.2 LA PLUPART DES PAGES AINSI FILTRES NE RENVOIENT PAS DE RESULTATS

Si la navigation par facettes est entièrement explorable, la majorité des filtres seront trop « fins » pour renvoyer des résultats.

Cela génère un grand nombre de pages « vides » ou présentant le message « zéro produit trouvé avec ces critères » (ou quelque chose du même style).

Notons que ces pages sont souvent interprétées par Google comme des « soft 404 » (apparaissant comme telles dans les « Outils pour Webmasters » alias les Google Webmaster Tools, voir la note en E.1.3.1), y compris et surtout si une règle de gestion redirige un visiteur qui atterrit sur une telle page vers la catégorie non filtrée...

C.1.3 NAVIGATION A FACETTES ET CONTENU DUPLIQUE

Même quand le filtre envoie des résultats, il arrive fréquemment que ces résultats soient identiques d'une page filtrée à une autre...

Imaginons que tous les T-shirt de la marque X soient en coton et de couleur blanc

Page filtrée selon marque = X : 10 produits

Page filtrée selon marque=X + matière = coton : 10 mêmes produits

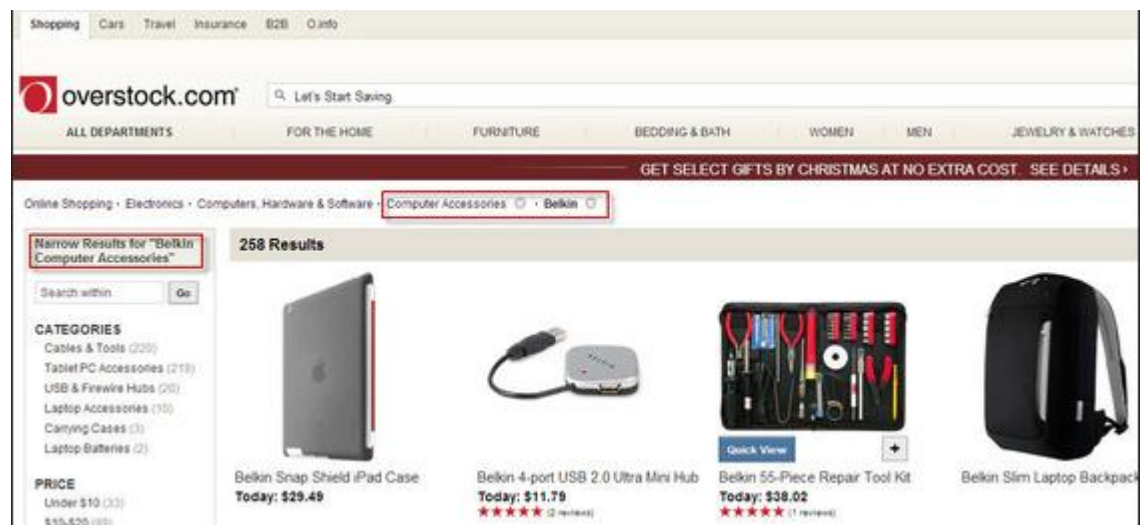
Page filtrée selon marque=X + couleur=blanc : 10 mêmes produits

Page filtrée selon marque=X + matière = coton + couleur = blanc : 10 mêmes produits...

Notons que certains filtres peuvent aussi générer des listings de produits presque identiques. (29 produits sur 30, 25 sur 30...)

C.1.4 LES FACETTES VUES PAR LES MOTEURS

Les robots d'exploration des moteurs de recherche voient vos pages sous un jour très différent de ce que vous montrez à l'utilisateur. Voici trois exemples de pages vues par les utilisateurs sur des sites e-commerce, présentant tous des produits de la marque Belkin :





Office DEPOT

FREE 1HR STORE PICKUP • FREE DELIVERY NO MINIMUM •

Search by Keyword or Item # Search

Weekly Ad

0 Item(s): \$0.00

Delivery Pick-Up View Cart Checkout

Welcome, Guest Log In | Register Browsing 60050 Change Zip Order Tracking

Office Supplies Paper Ink & Toner Breakroom Cleaning Technology Furniture Our Services My Account

Home > Technology > Computer & Tablet Accessories > Laptop Accessories

Laptop Accessories 8 Results Found

Shop Technology

- Laptop Cases
- Laptop Cooling Devices
- Laptop Security Cable Locks
- Laptop Stands & Lap Desks

Brands

- ☒ Belkin
- ☐ Targus
- ☐ Cocoon
- ☐ Case Logic
- ☐ Kensington
- ☐ 3M

See all Brands

Price

- ☐ \$10 and below
- ☐ \$10 - \$20





COMPARE: 0000 Compare up to 4 items

Sort By: Best Match

Showing 1-6 of 8 Results

Grid List

Items per page: 12

			
Compare	Compare	Compare	Compare
Belkin® CushTop Notebook Computer Cushion, Black Item # 303125	Belkin Notebook Security Key Lock Item # 154570	Belkin® Laptop Cooling Pad, Black Item # 760300	Belkin Master Key for Notebook Item # 150395

PC Connection

solving IT one customer at a time™

Search FIND

Small Business Business Enterprise Government & Education MacConnection

All Print page E-mail page Facebook Twitter Share Page

Accessories > Computer Accessories

1 - 20 of 269 Results

Results per page: 10 | 20 | 50

Sort By: Availability

Page: 1 2 3 4 5 6 7 8 9 10 | To View Your Pricing Log In

Compare	Product	Availability	Price
<input type="checkbox"/> Compare	Belkin Mousepad, Black Belkin Components • Item #: 303570 • Mfr #: [38009-BLK] • Platform: PC/Mac See All : Ergonomic Products	In Stock	\$1.95 ADD TO CART
<input type="checkbox"/> Compare	Belkin USB 2.0 4-Port Ultra Mini Hub Belkin Components • Item #: 7994933 • Mfr #: [5U407] • Platform: PC/Mac • 4 USB Ports [+] Show More Features See All : USB & Firewire Hubs	In Stock	\$6.95 ADD TO CART
<input type="checkbox"/> Compare	Belkin Mousepad, Standard Gray Belkin Components • Item #: 189577 • Mfr #: [38001-GRY] • Platform: PC/Mac See All : Ergonomic Products	In Stock	\$0.95 ADD TO CART

Et voici la version vue par les robots d'exploration

Overstock.com

Narrow Within "Belkin Computer Accessories"

Categories

- [Cables & Tools](#) (220)
- [Tablet PC Accessories](#) (219)
- [USB & Firewire Hubs](#) (20)
- [Laptop Accessories](#) (10)
- [Carrying Cases](#) (3)
- [Laptop Batteries](#) (2)

Price

- [Under \\$10](#) (33)
- [\\$10-\\$20](#) (89)
- [\\$20-\\$30](#) (69)
- [\\$30-\\$50](#) (52)
- [\\$50+](#) (15)

Product Types

- [Network Cable](#) (79)
- [Data Transfer Cable](#) (41)
- [A-V Cable](#) (19)
- [USB Hub](#) (18)
- [Video Cable](#) (15)
- [KVM Cable](#) (12)
- [Audio Cable](#) (8)
- [Analog KVM Switch](#) (5)
- [Cable Guide](#) (3)
- [Cable Lock](#) (3)
- [Carrying Case](#) (3)
- [Network Connector](#) (3)
- [Cable Panel](#) (2)
- [Cooling Stand](#) (2)
- [FireWire Hub](#) (2)
- [KVM Console](#) (2)
- [Network Patch Panel](#) (2)
- [USB Switch](#) (2)
- [Cable Cover](#) (1)
- [Case](#) (1)
- [Data Transfer Adapter](#) (1)
- [Handheld Device Cradle](#) (1)
- [Mouse Pad](#) (1)
- [Network Adapter](#) (1)
- [Repair Kit](#) (1)
- [Service Kit](#) (1)
- [Skin](#) (1)
- [USB Adapter](#) (1)

Office Depot

Laptop Accessories 8 Results Found

Shop Technology

- [Laptop Cases](#)
- [Laptop Cooling Devices](#)
- [Laptop Security Cable Locks](#)
- [Laptop Stands & Lap Desks](#)

Brands

- ☒ Belkin
- ☐ Targus
- ☐ Cocoon
- ☐ Case Logic
- ☐ Kensington
- ☐ 3M
- [See all Brands](#)

Price

- ☐ \$10 and below
- ☐ \$10 - \$20
- ☐ \$20 - \$30
- ☐ \$30 - \$40
- ☐ \$40 - \$50

Availability

- ☐ Sold Online
- ☐ Sold in Stores

Rating

- ☐ ★★★★★

Length

- ☐ 6 ft

Weight

- ☐ 1.00 lb

Color

PC Connection

Category
Computer Accessories

Manufacturer
Belkin Components

[Clear All](#)

Search within

Category

- [Ergonomic Products](#) (14)
- [Laptop Carrying Cases](#) (58)
- [Protective & Dust Covers](#) (145)
- [Tools & Hardware](#) (6)
- [USB & Firewire Hubs](#) (46)

Manufacturer

- [Belkin Components](#)
- [BTI - Battery Technologies](#) (645)
- [Targus](#) (411)
- [Axiom Memory Solutions](#) (368)
- [Mobile Edge](#) (299)
- [Protect Computer Products Inc](#) (215)
- [Denaq](#) (184)
- [Panasonic](#) (175)
- [Oncore Power Systems, Inc.](#) (167)
- [Fujitsu](#) (158)
- [Ereplacements, LLC](#) (117)

[More...](#)

List Price

- [Under \\$10](#) (22)
- [\\$10 - \\$25](#) (61)
- [\\$25 - \\$50](#) (144)
- [\\$50 - \\$100](#) (30)
- [\\$100 - \\$200](#) (12)

Platform

- [Both](#) (269)
- [MAC](#) (255)
- [PC](#) (214)

Ratings

- [5 Star](#) (3)
- [4 Star and above](#) (8)
- [3 Star and above](#) (8)
- [2 Star and above](#) (8)
- [1 Star and above](#) (9)

Color

- [Black](#) (88)
- [Blue](#) (14)
- [Brown](#) (2)
- [Burnt orange](#) (1)
- [Cabernet](#) (3)

Voici les « facettes » tel qu'un robot d'exploration comme Googlebot pourrait les voir. Chez Office Depot, les facettes ne sont pas explorables, à l'exception de l'attribut shop technology. Chez Overstock, certains les formulaires vus par l'utilisateur se transforment en liens.

C.1.5 MATT CUTTS ET LE CONTENU A FACETTES

Matt Cutts est le responsable de l'équipe « Qualité de Recherche » de Google, en charge de la lutte contre le webspam. Dans une interview accordée à Eric Enge le 14 mars 2010¹, il a rappelé les problèmes que posaient les facettes à Google dans la phase d'exploration.

Voici des extraits de ses commentaires :

Eric Enge: Parlons un peu de la navigation à facettes. Par exemple, sur Zappos, on peut acheter des chaussures filtrées par taille, par couleur, par marque, et le même produit est listé de 20 manières différentes, donc cela peut être très problématique. Que pensez-vous de ce genre de situations ?



Matt Cutts: Une navigation à facettes peut être très piègeuse, en règle générale.. Des utilisateurs réguliers ne la manipulent pas toujours bien, and ils se retrouvent un peu perdus. Il peut y avoir différentes façons de naviguer dans un contenu, mais vous allez préférer que chaque page de contenu ait une url unique si c'est possible.

Il y'a de multiples façons de morceler des données. Si vous pouvez déterminer vous-mêmes quelle est la meilleure manière d'obtenir un contenu en particulier, alors vous pouvez concrètement vous attaquer à la création d'une sorte de hiérarchie dans les paramètres.

La catégorie sera le premier paramètre, et le prix le second. Même si quelqu'un navigue d'abord via un filtre prix, puis clique sur une catégorie, vous pouvez faire en sorte que la hiérarchie que vous avez établie se reflète dans la position des paramètres.

De cette façon, la catégorie la plus importante apparaîtra en premier dans l'url, et la suivante, en second. Cette approche peut aider les moteurs de recherche à découvrir un peu mieux votre contenu, parce qu'ils peuvent se rendre compte que s'ils retirent ce dernier paramètre, ils obtiendront soit du contenu utile ou similaire.

En règle générale, la navigation à facettes est un problème sérieux, parce que vous pouvez créer un grand nombre de manières différentes d'accéder à une page. Vous pouvez avoir un nombre important de chemins intermédiaires à parcourir avant d'accéder au contenu utile.

¹ <http://www.stonetemple.com/articles/interview-matt-cutts-012510.shtml>



S'il est possible de limiter la profondeur et le nombre des pages intermédiaires, c'est potentiellement une bonne pratique. Si quelqu'un doit cliquer successivement sur 7 niveaux de navigation à facettes pour trouver un seul produit, il peut perdre patience. C'est un peu bizarre du point de vue des moteurs de recherche d'avoir à traverser 7 ou 8 couches de navigation à facettes avant de trouver un produit. D'une certaine façon, c'est beaucoup de clics, et de pagerank utilisés sur des pages intermédiaires sans produits que les gens puissent acheter. Chacun de ces clics est une opportunité pour qu'un petit pourcentage de pagerank soit perdu. »

D.Optimiser sa navigation à facettes

On vient de le voir, la grande majorité des combinaisons de facettes créent des pages sans intérêt pour l'utilisateur et sans intérêt, voire contre productives, pour le SEO.

Mais il existe a contrario des combinaisons de facettes qui :

- Créent des filtres utiles qui peuvent aider un utilisateur à trouver les items qu'il cherche dans un catalogue
- Et créer des pages d'atterrissage pertinentes (pour le SEO, comme pour le SEA)

D.1.1 TRIER LE BON GRAIN DE L'IVRAIE : IDENTIFIER LES « BONNES » FACETTES POUR LE SEO

Les facettes proposées sous forme de filtres actionnables par les utilisateurs sont choisies selon des critères « métiers » (et d'ergonomie)

Mais il faut garder en tête que les critères présidant à l'élection d'une facette au titre de « bon filtre pour le SEO » sont bien différents.

Les filtres intéressants sont ceux qui sont susceptibles de correspondre à des requêtes d'internautes. Par exemple, le filtre « couleur=rouge » appliqué à la catégorie « robe » crée une page d'atterrissage adaptée sur la requête robe rouge.

Mais toutes les facettes ne sont pas bonnes pour générer des pages d'atterrissage :

S'il est logique de chercher des « robes Desigual rouges », la requête « Carte SD rouge » est beaucoup plus improbable.

D.1.1.1 LA PERTINENCE D'UNE FACETTE OUVERTE AU SEO DEPEND DU CONTEXTE

Dans la pratique, on aura deux catégories de filtres :

- Des filtres génériques « transunivers / transcatégories », correspondant à des attributs associés à des items dans des univers/catégories différents : l'archétype de ce genre d'attribut c'est le prix, présent en général comme attribut universel pour tous les produits. Mais on peut citer aussi la marque.
- Des filtres spécifiques « attachés à des catégories » : ex la capacité pour les cartes mémoire..., la diagonale pour les écrans etc...

Pour les attributs « génériques », il peut sembler possible de créer une règle universelle. Par exemple, on peut constater que les requêtes incorporant la marque sont fréquentes, donc les filtres contenant la facette marque= « xxx » seront considérés comme pertinents.

Oui mais attention : si pour une catégorie, les marques sont inconnues des utilisateurs, le filtre de marque perd de son intérêt.

On peut faire le même genre de remarques pour le filtre prix :

- Les requêtes avec un prix précis sont rares => on peut « oublier » le filtre prix pour faire des pages ouvertes au SEO
- Sauf que les requêtes sur des bornes de prix sont moins rares (TV à moins de 500 euros)

On peut faire le même genre de remarques sur les filtres spécifiques à une ou plusieurs catégories : déterminer l'intérêt des filtres pour le SEO demande de tester le nombre de requêtes par mois effectuées correspondant à des listings d'items filtrées sur ces attributs, et de faire fonctionner son bon sens.

Mais en conclusion, l'intérêt d'une facette dépend toujours du contexte (un univers, une catégorie, la combinaison avec d'autres filtres) et l'ouverture ou la fermeture d'une facette doit se décider finement (au niveau « catégorie/sous-catégorie » au moins.

E. Comment ouvrir ou fermer une « facette »

E.1 *Gérer les différents états pour une facette*

En pratique, sur un site e-commerce, il est intéressant de pouvoir « administrer » les facettes, pour gérer de manière différenciée les filtres proposés aux utilisateurs, les pages créées à partir des facettes, et leur ouverture ou non au SEO

E.1.1 MISE A DISPOSITION DU FILTRE DANS L'INTERFACE UTILISATEUR

Dès lors que l'on dispose des informations dans la base de données catalogue permettant de créer un filtre de type « facette », il peut être tentant de proposer ce filtre aux utilisateurs. Dans la pratique, seul un petit nombre de facettes sera réellement utilisé dans l'interface, et ajouter trop d'options nuit à l'ergonomie des filtres.

Il est donc utile de disposer d'une interface permettant de « piloter » l'affichage du filtre dans l'interface utilisateur. C'est en particulier utile pour les filtres sur des attributs communs à différentes catégories ou univers : certaines facettes (ex : la matière) sont très utiles dans la catégorie « vêtements », moins dans la catégorie téléphone portable.

E.1.2 OUVERTURE DES PAGES FILTRÉES POUR LES MOTEURS DE RECHERCHE (REGLES A PRIORI)

Il est utile (et même indispensable) de pouvoir piloter l'ouverture des pages filtrées par facettes au SEO. Cela signifie dans la pratique gérer différents états

E.1.2.1 URL DE LISTING FILTRÉE PAR FACETTES FERMÉE AUX MOTEURS DE RECHERCHE

La syntaxe d'une url filtrée par facettes doit renvoyer 404 dans ce cas.

E.1.2.2 URL DE LISTING FILTRE PAR FACETTES OUVERTE A L'EXPLORATION D'UN MOTEUR DE RECHERCHE

La syntaxe d'une url filtrée par facette renvoie un code 200.

E.1.3 REGLES DE GESTION GEREES A POSTERIORI POUR LES URLS FILTRES PAR FACETTES OUVERTES AU SEO

Les règles a priori servent à indiquer qu'en principe, une facette va générer des pages que l'on souhaite voir crawler et indexer par les moteurs de recherche. Mais on peut constater que les pages ainsi générées présentent quelques défauts fâcheux :

E.1.3.1 LA PAGE FILTREE NE RENVOIE AUCUN RESULTAT

C'est une mauvaise idée d'avoir beaucoup de pages de ce type, que Google va identifier facilement comme des soft 404², et globalement donne un signal de mauvaise qualité quant au contenu du site)

Règles de gestion classiques **si cet état est définitif** :

- L'url doit renvoyer un code 404
- L'url disparaît du maillage interne (elle n'apparaît plus dans la page de catégorie, ou sur d'autres listings filtrés)
- La page est transformée en «page 404 intelligente», et affiche des items similaires, les catégories voisines etc...

Si cet état est temporaire (ex : rupture de stock, puis reconstitution du stock)

- L'url est placée en noindex, follow temporairement
- L'url disparaît du maillage interne (on ne propose pas un produit indisponible sur le catalogue)
- La page est transformée en «page sans résultats intelligente», et affiche des items similaires, les catégories voisines + une mention adaptée («plus d'articles disponibles : stock épuisé, en attente de réapprovisionnement» par exemple)

Ce type de pages est appelé communément en anglais « **smart no results page** ».

Remarque : il est formellement déconseillé de faire des redirections vers la catégorie parente, comportement souvent observé, mais qui a deux inconvénients majeurs :

- **C'est décevant pour l'internaute, qui s'attend à trouver un certain type d'items et tombe (sans explication le plus souvent) sur une page d'atterrissage plus générique : rebond assuré, et une conversion perdue**

² Google cherche à détecter les cas de pages introuvables pour lesquels le serveur web ou l'appliquatif web ne renvoie pas un code 404, comme ils le devraient, mais un code 200 (found !). Ce problème est incroyablement commun. Google appelle ces pages d'erreur qui ne renvoient pas de codes 404 des « soft 404 »



- Cela empêche la conservation des signaux historiques pour les cas de sortie temporaire, et pour les autres, cela peut générer un grand nombre de redirections assimilés à des soft404, avec des résultats potentiellement inattendus sur l'indexation et le positionnement des pages de destination de ces redirections

E.1.3.2 LA PAGE FILTREE RENVOIE TRES PEU DE RESULTATS (UN ITEM, DEUX ITEMS...)

Selon le contexte, ce sera soit acceptable, soit considéré comme une page de médiocre qualité.

Si page est considérée comme de médiocre qualité, on applique cette règle de gestion :

- L'url est placée en noindex, follow temporairement (jusqu'au retour à plus de résultats)
- La page est transformée en «page filtrée + requête élargie», et affiche des items complémentaire issues de filtres plus larges, et/ou des items similaires, des catégories voisines + une mention adaptée («indisponible : stock épuisé, en attente de réapprovisionnement » par exemple)

Tout l'art pour créer des pages efficaces consiste à savoir comment « élargir » la requête... Classiquement, on retire un filtre, et si possible, le filtre le plus sélectif en cas de filtres multiples.

Ce type de page est appelée communément en anglais « **smart few results page** » .

E.1.3.3 LA PAGE FILTREE AVEC UN JEU DE FACETTES RENVOIE LA MEME LISTE D'ITEMS QU'UNE AUTRE PAGE FILTRE AVEC UN AUTRE JEU DE FACETTES.

On a généré des quasi-doublons, qu'il est délicat de régler via des rel='canonical' car les pages sont potentiellement trop différentes.

Dans ce cas la règle de gestion est la suivante :

- On désigne (par une règle) quelle est la page qui doit rester dans l'index
- Les autres urls sont placées en noindex, follow
- On peut décider de garder toutes les pages en noindex dans le maillage ou non (règle métier propre à chaque site)

F. Navigation à facettes et performances

Créer des pages de listings filtrées avec des facettes sollicite potentiellement beaucoup les bases de données et/ou le moteur de recherche interne. Or les bases de données sont les points faibles des applicatifs web, et les moteurs de recherche sont souvent lents à renvoyer des résultats dès que la requête devient complexe.

Il est donc tout à fait habituel de constater des problèmes de performances avec ces pages :

- Serveurs très sollicités par la consultation de ces pages
- Temps de génération des pages beaucoup plus longs que la moyenne
- Timeouts / erreurs 500 fréquentes lorsque ces urls sont demandées

La conséquence immédiate, c'est que l'exploration de ces pages est rendue difficile pour les moteurs de recherche. Le crawl des moteurs est très sensible au temps de téléchargement des pages, qui se compose :

- Des temps de latence : en particulier temps pour que la requête du navigateur arrive au serveur web ou au proxy
- Du temps de génération de la page par un script
- Du temps de latence en retour : délai de réponse du serveur web pour renvoyer le premier octet de la page
- Du temps de téléchargement du code par les moteurs

F.1.1 OPTIMISER LES TEMPS DE TELECHARGEMENT DES PAGES FILTRES PAR FACETTES

Il est donc primordial de réfléchir à des solutions pour optimiser les temps de génération des pages (c'est un problème général mais qui devient aigu pour les pages filtrées par facettes).

De nombreuses pistes d'optimisation existent :

- Optimisation des bases et des index
- Caches niveau code, caches requêtes, caches HTML
- Reverse proxies
- Utilisation de CDN et de serveurs « edge » (Akamai, Limelight, ...)

G. Exploiter les facettes dans un moteur de recherche interne

Certains moteurs de recherche internes (SolR par exemple) proposent une fonctionnalité qui améliore considérablement l'expérience utilisateur : le renvoi direct vers la page d'atterrissage appropriée !

G.1 *Comment cela fonctionne-t'il ?*

Si un utilisateur tape une requête du type « peugeot 306 motorisation essence bleue », le moteur de recherche interne identifie les termes composant la requête de la manière suivante :

- Peugeot -> marque, filtre : marque = peugeot
- 306 -> modèle, filtre : modele = 306
- Motorisation essence -> attribut+ valeur, filtre : motor=essence
- Bleue -> couleur de carrosserie, filtre : peinture = bleue

Ou tout simplement, propose l'url la plus proche de ces termes et sélectionne l'url suivante

<http://www.domaine.com/listeannonces.php?marque=peugeot&modele=306&motor=essence&peinture=bleue>

Le résultat, c'est que le moteur interne reroute l'utilisateur, non pas vers une page de recherche, mais vers une page de catégorie filtrée.

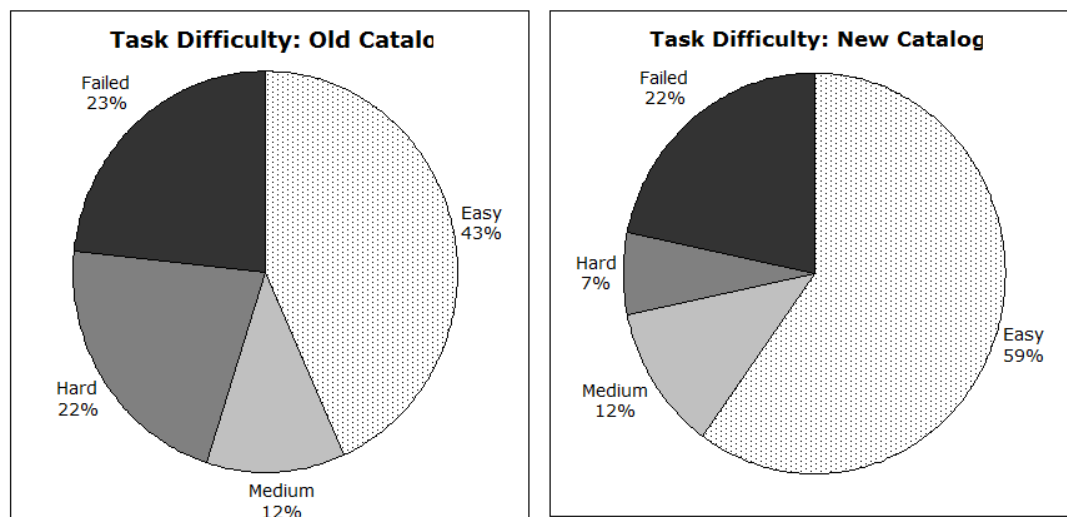
G.2 Les avantages

Si la redirection est permanente, le gain en terme de conversion est évident : la page de catégorie filtrée est mieux conçue pour transformer qu'une page de résultats de recherche classique.

Il y'a un deuxième avantage pour le SEO : envoyer du trafic vers des urls qui sinon ne seraient visitées que par des utilisateurs en provenance des moteurs de recherche est une bonne pratique.

H. Quels gains espérer de ces optimisations

H.1 Gain n°1 : amélioration de l'expérience utilisateur



Une étude d'ergonomie réalisée par l'université de Californie sur le site de sa bibliothèque : la recherche a été estimée « facile » dans 59% des cas avec la navigation à facettes contre 43% des cas avant, et le nombre de cas « difficiles » est tombé de 22% à 7%

Comme dans l'étude ci-dessus, on peut démontrer que l'ajout d'une navigation à facettes apporte une amélioration de l'expérience utilisateur, et ce, quel que soit le contexte (site e-commerce, recherche documentaire, autres fonctionnalités de recherche).

La recherche par mots clés peut être efficace, à condition de savoir s'en servir : beaucoup d'utilisateurs perdent du temps à « essayer » différents mots clés sans jamais identifier ceux qui filtrent les résultats de manière efficace. Par ailleurs, elle nécessite que l'utilisateur saisisse ces termes dans un champ, ce qui requiert l'abandon du dispositif de pointage et le recours à un clavier parfois absent (nouveaux dispositifs comme les tablettes) ou inadaptés (certains smartphones)

La recherche par liens présente l'avantage d'une navigation ultra simple pour les utilisateurs paresseux qui peuvent parcourir un site uniquement à l'aide d'un dispositif de pointage (souris). Mais une navigation entièrement basée sur les liens demande à l'utilisateur de cliquer un grand nombre de fois avant d'accéder à ce qu'il recherche.

La recherche par facettes permet de réduire le nombre de clics nécessaires pour filtrer efficacement le résultat et trouver les items recherchés, tout en affichant la liste des filtres disponibles.

H.2 *Gain n°2 : plus de trafic capté*

L'exploitation des facettes permet de créer un grand nombre de pages de catégories filtrées par facette, qui constituent autant d'excellentes pages d'atterrissage sur des requêtes de type « catégorie filtre1 valeur1 filtre2 valeur 2... »

Ces pages d'atterrissage peuvent servir :

- Pour se positionner sur ces requêtes en SEO
- Comme pages d'atterrissage adwords (avec potentiellement, une meilleure conversion, et un meilleur quality score que des pages trop génériques, ce qui ne gâte rien)
- En affiliation
- Et même pour des campagnes display !

Le gain de trafic apporté par ces nouvelles pages d'atterrissage se révèle souvent significatif.

H.3 *Gain n°3 : plus de conversions*

On démontre souvent (par exemple lors de testing A/B) que ces pages performant mieux que d'autres types de pages.

En particulier, ces pages convertissent mieux qu'une page de recherche « fulltext » sur les mots clés liés aux facettes.



Et elles convertissent mieux que des pages plus génériques.

I. Ce que Search Foresight peut faire pour vous aider

Grâce à leurs missions passées, les consultants de Search Foresight ont acquis une expérience unique sur des technologies régulièrement employées pour créer des navigations à facettes : Oracle Endeca, Oracle ATG, Lucène, SolR, Elastic Search, Intuiko, Exalead, Antidot, HP Autonomy, mais aussi de nombreuses plateformes e-commerce comme Hybris, Intuiko, Magento, Websphere...

Nous pouvons donc aider vos équipes à implémenter nos recommandations, en intervenant en particulier sur :

- La **récolte d'information structurées** via le data mining ou des bases de données métiers externes
- Le **stockage de l'information** et la **création de nouvelles facettes**
- **L'identification des facettes utiles pour l'utilisateur, et/ou pour le SEO**
- L'implémentation des **systèmes de pilotage d'ouverture/fermeture des facettes** au crawl, à l'indexation, aux utilisateurs...
- L'implémentation d'intégrations HTML adaptées pour créer à la fois des **IHM (interfaces utilisateur) riches, ergonomiques et compatibles SEO**
- La mise en place de solutions de **gestion dynamique des pages ouvertes** en fonction de l'évolution du catalogue :
 - Entrée / sortie de produits
 - Stock épuisés, produits indisponibles
 - Pages : « Smart no results », « Smart few results », « Smart 404 »

- Gestion des doublons et des quasi doublons
- Nous pouvons aussi former vos équipes à ces techniques grâce aux [modules de formation](#) proposés par le [Centre de Formation de Search Foresight](#)

J. Conclusion : structurez toutes vos informations pour mieux les exploiter

Maîtriser sa navigation à facettes permet d'en exploiter ses nombreux avantages pour le SEO : c'est-à-dire, créer des pages d'atterrissage efficaces, avec un taux de conversion correct, et répondant à un grand nombre de requêtes effectuées par les internautes sur les moteurs de recherche.

Tout l'art consiste ensuite à éviter les effets de bords : la création de doublons, de pages vides ou sans intérêt, et un ratio trop élevé entre pages de listings et fiches produit.

Il est important également d'adapter son maillage interne pour optimiser son site dans ce nouveau contexte, qui crée potentiellement un grand nombre de nouvelles pages. Le nouveau maillage devra tenir compte de la nouvelle hiérarchie d'importance entre les anciennes pages et les nouvelles, et respecter les règles habituelles : siloisation, enrichissement sémantique des pages, optimisation de la diffusion du « linkjuice ».

Mais la navigation à facettes suppose que l'on puisse identifier facilement dans votre contenu des « attributs » et les valeurs de ces attributs. Si l'essentiel des caractéristiques produit ne sont pas enregistrés dans votre PIM (Product Information Manager) ou dans la base de données de votre plateforme e-commerce sous forme de champs structurés, mais juste mentionnée dans un champ texte de type descriptif produit, vous avez un problème ! Dans ce cas, pas de recherche facettée possible, vous en serez réduit à une recherche dite « full text », dont les résultats sont souvent de piètre qualité.

Une des tâches prioritaires consiste donc à « structurer » l'information, c'est-à-dire de stocker dans des champs structurés les couples attributs valeurs. Deux méthodes sont couramment utilisées :

- Soit enrichir la base de produits à l'aide de bases de caractéristiques produit, fournies sous forme structurée par le fabricant (sous forme de fichiers xml par exemple), soit par des sociétés spécialisées
- Soit, si les seules sources d'information sur les caractéristiques produit, utiliser des techniques sémantiques avancées pour annoter et étiqueter les produits en identifiant de manière automatique les attributs mentionnés dans des descriptions.

Ces technologies sémantiques, exploitables dans un site e-commerce, font partie d'une famille de techniques de type « data mining », baptisée en anglais « information extraction » parmi lesquelles on trouve la « reconnaissance des entités nommées », et « l'extraction automatique de relations ».

Mais cela fera l'objet d'un prochain Livre Blanc de Search Foresight...



SEARCH • FORESIGHT