

Les outils de recherche d'images (3ème partie)

Nous continuons notre série d'articles sur les outils de recherche spécialisés sur certains formats spécifiques de fichiers ou de données : image, actualité, fichiers PDF, etc. Les mois précédents, nous avons étudié les moteurs de recherche d'images avec la fonction "Images" des outils de recherche traditionnels (Google, Yahoo!, Lycos, etc.) et les outils spécialisés (Ditto, PicSearch, Corbis).

Ce mois-ci, nous étudierons les métamoteurs d'images. Ils interrogent les moteurs d'images existant et en synthétisent les résultats, tout comme un métamoteur web "classique". Sont-ils plus précis, plus pertinents, que ceux que nous avons décrits les mois précédents (comparatifs dont Google est sorti vainqueur) ? Pour le savoir, nous leur avons appliqué le même mode de comparatif quantitatif, qualitatif et fonctionnel.

Nous avons sélectionné les métamoteurs de recherche d'image suivants :

- Ithaki (<http://images.ithaki.net/>). Effectue ses recherches sur Google/Ubbi, Dino Cobion, Visoo, Altavista Images, Fast Picture Search/All the web, Picsearch et Art.com. Ithaki est le seul métamoteur réellement spécialisé dans les images trouvées sur le Web. Les autres métamoteurs, ci-dessous, proposent un choix "Images", mais également des recherches sur le Web ou d'autres formats (audio, etc.).
- Mamma (<http://www.mamma.com/>). Effectue ses recherches sur Lycos Images, WebShot et PicSearch.
- Fazzle (<http://www.fazzle.com/>). Effectue ses recherches sur AltaVista Images, Lycos Images, WebShot, Fast Multimedia et PicSearch.
- Ixquick (<http://www.ixquick.com/>). Effectue ses recherches sur Fast Multimedia, GoGraph et PicSearch.
- MetaHoo (<http://www.metahoo.com/>). Effectue ses recherches sur Fast Multimedia, AltaVista, WebShot et Google.

Un outil comme DogPile (<http://www.dogpile.com/> puis option "Images" sur les pages de résultats) fournit les résultats de Fast puis ceux de Ditto, sans synthèse, tout comme Excite (<http://www.excite.com/>) et WebCrawler (<http://www.webcrawler.com/>), moteur du même groupe (Infospace). L'intérêt nous semble, du coup, assez faible.

Comparatif quantitatif

Dans un premier temps, nous avons tapé les mêmes mots clés que les mois précédents (en anglais, la plupart des sites testés étant d'origine anglophone et proposant des résultats assez pauvres sur des termes francophones) sur chaque outil de recherche et avons noté le nombre de résultats proposé par chaque outil :

Moteur :						
napoleon	46	19	161	19	45	11 400
strasbourg cathedral	49	19	121	17	41	397
butterfly	47	19	152	28	56	177 000
madonna	49	19	136	18	47	57 400
ferrari	57	19	114	18	41	108 000
bush	49	19	123	27	56	178 000
ibm logo	46	19	87	3	38	6 690
key	47	19	100	27	49	316 000

tatooo	26	19	134	16	29	6 590
harry potter	47	19	121	16	42	87 100
lance armstrong	46	19	118	15	41	4 550
mona lisa	46	19	140	11	39	5 220
Moyenne :	0,01	0,00	0,03	0,00	0,01	20,00

Tous les métamoteurs d'images semblent limiter le nombre d'images proposées dans leurs résultats. Ithaki n'a jamais indiqué plus de 57 images, même sur des thèmes très "porteurs" (Harry Potter, Madonna, Ferrari, etc.). Pour Mamma, la limite "haute" semble clairement avoir été mise à 19 images... Fuzzle est plus explicite : "Votre recherche pour bush a retourné 585,600 résultats. Nous avons choisi les premiers 123." Etc. Cette limitation est-elle due à des accords avec les moteurs interrogés ? Et-ce une limitation pour ne pas connaître de problèmes de bande passante ? Dans tous les cas, difficile, alors, de concurrencer Google au niveau quantitatif...

Pour donner une note à chaque moteur, nous avons attribué, pour une requête donnée, la note de 20 au moteur donnant le plus grand nombre de résultats. Puis, une règle de trois nous a permis de calculer les notes des autres moteurs.

Comparatif qualitatif

Pour comparer de façon qualitative les 5 outils étudiés, nous avons tapé les mêmes 12 requêtes sur chacun d'eux. Puis, nous avons évalué les 20 premières réponses (les 20 premières images affichées) et noté combien, parmi elles, étaient pertinentes. Comme les mois précédents, la notion de pertinence pouvait être assez générale ou plutôt spécifique, en fonction du mot clé. Par exemple, nous avons considéré comme pertinentes, pour le mot clé "bush", des photos de Georges Bush père et fils, de la chanteuse Kate Bush, ou de toute personne portant ce nom. En revanche, pour le mot clé "butterfly", nous n'avons pas accepté, par exemple, les photos du chien ou du chat portant ce nom, car nous estimons que le moteur doit être capable de remettre la requête dans son contexte et nous proposer une photo représentant, de façon globale, des papillons, thème certainement plus "populaire" que les chiens portant ce nom. Nous avons parfois tenu compte d'un parti pris, comme pour Madonna, mot clé pour lequel nous n'avons noté que les images représentant la chanteuse. Mais ce même parti-pris a été maintenu sur tous les outils.

Moteur :						
napoleon	12	13	9	9	16	20
strasbourg cathedral	16	13	16	12	17	18
butterfly	14	12	14	12	17	12
madonna	13	18	10	1	7	20
ferrari	20	16	20	18	20	20
bush	10	5	6	4	7	19
ibm logo	17	11	13	7	17	17
key	7	3	6	3	11	2
tatooo	15	15	13	9	17	19
harry potter	20	10	10	14	17	18
lance armstrong	20	20	19	16	20	19

mona lisa	18	13	14	8	19	16
Moyenne :	15,17	12,42	12,50	9,42	15,42	16,67

A noter que Mamma semble avoir un bug : lorsqu'on tape une expression sur plusieurs mots clés ("mona lisa"), la première page de résultats est exacte, mais si on clique sur le lien "Next", la deuxième page de liens semble avoir perdu le deuxième mot clé (la recherche s'effectue, alors, sur "Mona" uniquement).

Sur Fazzle, nous n'avons pas tenu compte des premiers liens proposés, qui correspondent à des publicités. Notons également que, sur Fazzle, les vignettes représentant les images sont tellement petites qu'il est souvent complexe de voir ce qu'elles représentent... Heureusement, un lien "Preview" affiche une frame contenant l'image et plus d'infos à son sujet.

Sur le plan qualitatif, Ithaki et MetaHoo rivalisent avec Google. Fazzle et Mamma sont un peu en retrait mais restent de bons outils de recherche. Ixquick, en revanche, semble distancé pour ce qui est de la qualité des résultats retournés.

Comparatif fonctionnel

Nous avons noté ici, les facilités de recherche (syntaxe spécifique, présence d'une zone de recherche avancée, de filtres, etc.) ainsi que les informations fournies pour chaque image.

Moteur :						
Syntaxe avancée	Pas de syntaxe avancée disponible	AND, OR, NOT, guillemets	Et, OU, phrase, titre, booléen	ET, OU, SAUF, parenthèses, guillemets, etc.	Pas de syntaxe avancée disponible	OR, -, guillemets, filetype:
Recherche avancée / filtres	Nombre de résultats par page, présentation plus ou moins compacte, filtres géographiques	Source, nombre de résultats par page, filtre familial, signalement des termes recherchés, filtre sur nom de domaine, etc.	Source, nombre de résultats par page, signalement des termes recherchés, etc.	Pas de recherche avancée disponible	Pas de recherche avancée disponible	Taille, format, couleur, domaine, filtre parental (US)
Informations fournies	Taille, poids, source, lien	Source	Titre, taille, lien, lien "Preview" qui affiche une frame contenant l'image et plus d'infos à son sujet.	Titre, taille, poids, source	Titre, taille, poids, source	nom, taille, poids, url
Note :	5	15	17	12	5	15

Récapitulatif

Moteur :						
Quantitatif	0,01	0,00	0,03	0,00	0,01	20
Qualitatif	15,17	12,42	12,5	9,42	15,42	16,67
Fonctionnel	5	15	17	12	5	15
Moyenne :	8,84	9,96	10,51	7,71	8,96	17,09
Les "Plus"	Bons résultats qualitatifs				Bons résultats qualitatifs	Bonne pertinence moyenne, complémentarité avec d'autres outils de recherche (web, forums, etc.)
Les "Moins"	Faible nombre de résultats indiqués. Pas de syntaxe avancée.	Faible nombre de résultats indiqués	Faible nombre de résultats indiqués	Faible nombre de résultats indiqués (mais celui qui en fournit le plus parmi ses challengers). Pas de recherche avancée. Résultats qualitatifs moyens.	Faible nombre de résultats indiqués. Pas de syntaxe ni de recherche avancée.	

Pour calculer la note finale, nous avons donné un coefficient 1 à la note des comparatif quantitatif et fonctionnel et un coefficient 2 à la note du comparatif qualitatif, car celui-ci nous semble le plus important.

Le résultat donne, une fois de plus, Google gagnant haut la main. Si certains métamoteurs, notamment Ithaki et MetaHoo, sont intéressants qualitativement, ils pèchent aujourd'hui par un trop faible nombre de résultats fournis et surtout par des possibilités de recherche avancée trop faibles...

Bref, selon nos critères, Google a passé toutes les épreuves et peut être considéré aujourd'hui comme le meilleur outil de recherche d'images, toutes catégories confondues...