

Les outils de recherche d'images (2ème 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. Le mois dernier, nous avons débuté l'étude des moteurs de recherche d'images avec la fonction "Images" des outils de recherche traditionnels (Google, Yahoo!, Lycos, etc.).

Ce mois-ci, nous étudierons plus particulièrement les banques de données d'images et moteurs spécialisés. En effet, depuis plusieurs années, de nombreux sites se sont créés spécifiquement pour proposer ce type de recherche. Ce ne sont pas des moteurs de recherche sur le Web mais uniquement des moteurs d'images spécialisés dans ce domaine. Sont-ils plus précis, plus pertinents, que ceux que nous avons décrits le mois dernier (comparatif dont Google est sorti vainqueur) ? Pour le savoir, nous leur avons appliqué le même mode de comparatif quantitatif, qualitatif et fonctionnel.

Ce mois-ci, nous avons donc testé 3 moteurs de recherche d'images, de loin les plus connus de ce domaine :

- Ditto (<http://www.ditto.com/>)
- PicSearch (<http://www.picsearch.com/>)
- Corbis (<http://www.corbis.com/>)

Ditto est un moteur américain, présent sur le marché depuis de nombreuses années. Il s'appelait auparavant Arriba. Son index est créé sur une double base humaine et automatique. Ditto, encore sous le nom d'Arriba, avait été mis sous les feux de la justice en 1999 (puis encore en 2002), notamment au travers de procès intenté par Leslie A. Kelly, photographe et éditeur américain, contestant les méthodes du moteur pour indexer les images (voir lettre R&R du mois de mars 2003 et son article intitulé "Législation : la jurisprudence autour des outils de recherche").

PicSearch est l'un des moteurs de recherche dédié à l'image les plus connus du Web. Il est d'origine suédoise.

Corbis est l'une des plus sociétés les plus connues du monde de l'image numérique et de la photographie. Elle appartient à Microsoft et propose deux types de recherches : pour les professionnels et pour le Grand public. De nombreuses images parmi celles qu'il propose ne sont pas issues de pages web mais de collection disponibles en ligne, libres de droit ou non.

Il existe cependant de très nombreux autres outils en ligne, comme :

- Art.com (<http://www.art.com/>), plutôt spécialisé dans les images d'art.
- Le moteur de recherche d'images du CNRS (<http://phototheque.cnrs.fr/>).
- GoGraph (<http://www.gograph.com/>), plutôt spécialisé dans les graphiques.
- GoodShoot (<http://www.goodshoot.fr/>) qui propose, in fine, peu de résultats.
- GettyImages (<http://creative.gettyimages.com/source/home/home.asp>) qui propose, lui aussi, assez peu de résultats.
- Visoo - Cobion (<http://www.visoo.com/>), assez connu mais qui semble ne plus fonctionner.

Nous avons testé tous ces outils. Mais ils nous ont rapidement semblé plus spécifiques, thématiques ou moins fonctionnels, aussi nous avons limité notre comparatif aux 3 outils pré-cités. De nombreux autres, si vous ne trouvez pas votre bonheur dans ces pages, vous sont notamment proposés sur le site d'Indicateur (<http://www.indicateur.com/thematique/images.shtml>).

Nous n'avons donc gardé, pour réaliser notre comparatif, que 3 sites, qui nous ont semblé les meilleurs de leur catégorie. Nous avons de plus comparé leurs résultats à ceux de Google, vainqueur de notre test du mois dernier (<http://images.google.com/>).

Comparatif quantitatif

Dans un premier temps, nous avons tapé les mêmes mots clés que le mois dernier (mais 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 | 391 | 3 509 | 1 000 | 11 400 |
| strasbourg cathedral | 1 668 | 26 | 143 | 397 |
| butterfly | 2 411 | 27 141 | 1 000 | 177 000 |
| madonna | 1 270 | 11 809 | 1 000 | 57 400 |
| ferrari | 5 376 | 15 688 | 1 000 | 108 000 |
| bush | 2 372 | 24 996 | 1 000 | 178 000 |
| ibm logo | 10 521 | 128 | 2 | 6 690 |
| key | 4 322 | 52 669 | 1 000 | 316 000 |
| tatoo | 353 | 1 006 | 93 | 6 590 |
| harry potter | 4 611 | 6 528 | 276 | 87 100 |
| lance armstrong | 1 160 | 456 | 888 | 4 550 |
| mona lisa | 2 718 | 1 082 | 84 | 5 220 |
| Moyenne : | 5,05 | 2,80 | 0,74 | 18,12 |

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.

Remarque importante : Corbis limite son nombre maximal d'images proposées à 1 000. Il est impossible de savoir le nombre exact d'images identifiées avec les mots clés saisis s'il est supérieur à ce nombre. Ce qui explique sa faible note à ce comparatif, PicSearch proposant, pour sa part parfois plus de 52 000 images... Quant à Google, il semble difficile à battre dans ce domaine.

Si Ditto obtient le meilleur score des moteurs "dédiés", c'est surtout grâce à deux "top scores" sur les expressions "strasbourg cathedral" et "ibm logo", sans que l'on sache vraiment si, pour des expressions à deux termes, le moteur place un ET ou un OU entre les deux (aucune indication n'est fournie sur le site à ce sujet)... PicSearch semble pourtant, globalement, meilleur quantitativement, que son rival.

Ceci dit, Google semble rester clairement incontournable quant au choix quantitatif d'images qu'il propose lors des recherches...

Comparatif qualitatif

Pour comparer de façon qualitative les 3 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 le mois dernier, 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 | 15 | 9 | 20 | 20 |
| strasbourg cathedral | 11 | 15 | 18 | 18 |
| butterfly | 17 | 0 | 12 | 12 |
| madonna | 9 | 15 | 20 | 20 |
| ferrari | 18 | 20 | 18 | 20 |
| bush | 19 | 9 | 0 | 19 |
| ibm logo | 1 | 20 | 2 | 17 |
| key | 3 | 0 | 17 | 2 |
| tatoo | 15 | 17 | 16 | 19 |
| harry potter | 19 | 20 | 20 | 18 |
| lance armstrong | 15 | 20 | 20 | 19 |
| mona lisa | 14 | 16 | 18 | 16 |
| Moyenne : | 13,00 | 13,42 | 15,08 | 16,67 |

Dans ce domaine, c'est encore Google qui obtient la meilleure note, même si les résultats de Corbis sont excellents et très proches de ceux du moteur leader. PicSearch et surtout Ditto paient leur manque de régularité, certaines recherches donnant de très bons résultats, d'autres très mauvais.

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 | NON | ET (+), SAUF (-, NOT) | ET, OU, SAUF, guillemets, parenthèses, etc. | OR, -, guillemets, filetype: |
| Recherche avancée / filtres | NON | Choix Images/ animations, Couleur / N&B, Taille du fichier | Images libres de droits ou non, couleur, orientation, photographe, numéro d'image, nombre de personnages, etc. | Taille, format, couleur, domaine, filtre parental (US) |
| Informations fournies | Lien, taille | Lien "info" : Lien, taille, dimensions, nombre de couleurs, type de fichier, etc. | Lien "info" : nom du photographe, date, autres mots clés, etc. | nom, taille, poids, url |
| Note : | 5 | 15 | 20 | 15 |

Ditto ne propose aucune aide en ligne (le lien semble inactif à l'adresse <http://www.ditto.com/services.asp>), aucune syntaxe avancée, bref aucune facilité de recherche. Très décevant pour un moteur de ce niveau. Les informations fournies pour chaque image sont également sybillines : le lien de la source et le poids. Peut (largement) mieux faire...

PicSearch, en revanche, se singularise par son lien "info" proposé pour chaque image et qui affiche de très nombreuses informations très utiles tout en remplaçant chaque photo dans son contexte sur la page web d'origine.

Quant à Corbis, sa recherche avancée est exemplaire : à voir absolument pour comprendre comment une recherche d'images peut être très complète et efficace. Un "must" ! Bien sûr, son catalogue est certainement plus "balisé" que celui de Google puisque ne provenant pas obligatoirement que du Web, mais l'outil est quand même superbe sur ce point de vue. Un véritable outil de professionnel.

Nous avons en revanche baissé la note de Google par rapport à celle du mois dernier pour tenir compte de la nouvelle concurrence de Corbis, de loin le meilleur acteur au niveau fonctionnel.

Récapitulatif

| Moteur : |  |  |  |  |
|--------------------|---|---|--|---|
| Quantitatif | 5,05 | 2,80 | 0,74 | 18,12 |
| Qualitatif | 13 | 13,42 | 15,08 | 16,67 |
| Fonctionnel | 5 | 15 | 20 | 15 |
| Moyenne : | 9,01 | 11,16 | 12,73 | 16,62 |
| Les "Plus" | | Bonne moyenne en règle générale | La syntaxe d'interrogation et la recherche avancée très complètes | Bonne pertinence moyenne, complémentarité avec d'autres outils de recherche (web, forums, etc.) |
| Les "Moins" | Pauvreté des possibilités de syntaxe avancée | Résultats assez inégaux | Seulement 1 000 résultats maximum affichés | |

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.

Google sort donc encore une fois vainqueur de ce comparatif général. On se prend à rêver d'un moteur comme celui que propose Google avec les outils de recherche avancée de Corbis...

Le mois prochain, nous étudierons la troisième famille d'outils de recherche d'images : les métamoteurs, qui sont également nombreux sur le Web. Encore une fois, nous comparerons ces outils à Google qui reste pour l'instant seul maître à bord...