

Audit de site : www.cite-sciences.fr

Une nouvelle rubrique dans la lettre R&R, au sein de laquelle nous allons étudier le taux d'optimisation d'un site web par rapport aux critères de pertinence des moteurs de recherche. Ce mois-ci : le site de la Cité des Sciences à Paris.

Nous vous proposons, dans cette nouvelle rubrique, un audit de site web qui nous permet d'estimer la façon dont les principaux critères de pertinence des moteurs de recherche sont pris en compte par un site web. Nous ce cadre, nous allons "auditer" un certain nombre de sites dans les mois qui viennent. Il nous semble cependant important de préciser un certain nombre de choses :

- Nous essaierons de prendre en compte des sites de contenu, ayant un réel intérêt à obtenir un bon référencement à moyen et long terme, plutôt que des sites d'actualité "chaude". Bref, plutôt le site de la Cité des Science que celui de TF1...
- Il ne s'agit absolument pas de critiquer le site et ses concepteurs, mais plutôt d'illustrer nos propos au sujet de l'optimisation des sites web au travers d'exemples concrets. Dans ce cadre, le plus important pour vous n'est pas obligatoirement ce qui est dit dans l'article au sujet du site audité, mais plutôt les informations que vous pouvez en déduire pour le vôtre...
- Les sites web que nous allons auditer font peut-être l'objet d'un référencement professionnel par des sociétés spécialisées à l'aide pages satellites ou autres solutions techniques, ce qui pourrait pallier certains critères bloquants relevés dans l'article. Nous n'en tiendrons pas compte pour nous focaliser uniquement sur l'aspect "optimisation des pages existantes".
- Enfin, si vous désirez que votre site fasse l'objet d'un audit dans cette rubrique, contactez-nous.

Pour ce premier article, nous allons prendre pour exemple le site de la Cité des Sciences et de L'industrie de Paris, disponible à l'adresse <http://www.cite-sciences.fr/>. Passons en revue les différentes "zones chaudes" de ses pages pour le référencement.

1. Redirection

Premier point important : lorsque vous saisissez l'adresse <http://www.cite-sciences.fr/>, vous êtes automatiquement redirigé sur une autre page, selon votre configuration. Exemple : <http://www.cite-sciences.fr/francais/indexFLASH.htm>

Il s'agit d'une redirection classique, notamment vers une page utilisant la technologie Flash si vous avez le "plug-in" correspondant sur votre ordinateur. Le problème, pour le référencement, est que les moteurs de recherche et leurs robots n'aiment pas trop ce type de redirection automatique...

Pour en savoir plus, il faut aller voir le code de la page <http://www.cite-sciences.fr/>. Mais il s'agit de la page contenant la redirection, difficile à "attraper au vol". Pour en lire plus facilement le code HTML, voici une "astuce" simple :

- Allez sur le site Outiref (<http://ww.outiref.com/>)
 - Tapez l'adresse du site web à étudier. Ici : <http://www.cite-sciences.fr/>
 - Sur la page de résultat, vous avez, en titre la phrase "**Analyse de l'URL : <http://www.cite-sciences.fr/> (voir code source)**". Cliquez sur "**Voir code source**".
- Vous pouvez ainsi analyser le code source de cette page de redirection, voire le copier/coller dans votre éditeur HTML habituel sans problème.

Ce document contient effectivement une redirection automatique en JavaScript vers différentes pages d'accueil, en fonction de la présence ou non du plug-in Flash sur la machine du visiteur.

En revanche, les moteurs ne suivront pas les liens vers ces pages-cibles car la plupart des robots ne suivent pas les liens inclus dans les JavaScripts. Du coup, si rien n'avait été fait, les robots se seraient certainement arrêtés sur cette page d'accueil, n'auraient pu aller plus loin et n'auraient indexé que ce document pour le site entier. Dommage....

Heureusement, le webmaster a bien fait les choses :-). En effet, la page index.html du site n'est pas une "simple" page de redirection mais un document à part entière, avec titre, balises meta, texte, etc. Et surtout des liens vers la plupart des zones importantes du site.

En effet, comme nous l'avons dit, le robot des moteurs va s'arrêter sur cette page sans suivre les redirections proposées dans le JavaScript. Mais il va quand même indexer le contenu de cette page : Titre, balises meta, texte, etc. Et il va suivre les liens qui y sont proposés ! Le tour est joué ! Le robot accède ainsi aux pages internes du site sans passer par les sommaires spécifiques, au format Flash ou non. Bien entendu, le lien vers la "vraie" page d'accueil renvoie le robot vers la page d'accueil "sans Flash", cette technologie étant peu ou pas du tout prise en compte par les moteurs. Le problème du Flash est donc résolu grâce à une interface "Light", qui sera bien utile pour le référencement, présente à l'adresse :
<http://www.cite-sciences.fr/francais/indexLIGHT.htm>

Il s'agit ici d'un type de page satellite (que nous appellerons "page de redirection" dans la suite de l'article) créée pour des raisons techniques, à savoir la présence d'une version Flash (ou non) du site. Il s'agit ici d'un bon exemple de façon de pallier un problème technique et un critère bloquant pour les moteurs en utilisant ce type de page...

2. Titres des pages

Les titres des pages sont descriptifs et assez courts :

- Page de redirection : "**Cite des Sciences et de l'Industrie - Paris**"
- Page d'accueil Flash (et non-Flash) : "**Bienvenue à la Cité des Sciences et de l'Industrie**"

Les deux titres contiennent entre 7 et 10 mots. Ils sont donc dans "la norme". Ils affichent des mots clés importants (excepté "Bienvenue"). Ils sont globalement bien optimisés.

Les pages internes du site sont ensuite bâties différemment : un système de frames 100% (voir par exemple le code de la page http://www.cite-sciences.fr/francais/ala_cite/cite_pra/info_pra/global_fs.htm) affiche des pages web de contenu. Chaque page "mère" a le même titre : "Cite des Sciences et de l'Industrie". Chaque page "fille" a son titre spécifique : "A la Cité", etc. Parfois, cependant, les titres des pages filles auraient mérité plus de précision. Exemple : pour la page présentant l'exposition "Explora", le titre est "Bienvenue à la Cité des Sciences et de l'Industrie". Pour la page sur les handicapés, on a pour titre "Cité des Sciences et de l'Industrie".

Rappelons que, idéalement lors de l'utilisation de frames, chaque page, chaque document (pages "mères" et pages "filles") doivent être optimisés au niveau du titre, des balises Meta, etc.

En revanche, sur le site de la Cité des Sciences, les pages filles ne contiennent pas le code javascript de reconstitution de la page mère si on essaie d'afficher la page de contenu, ce qui est un handicap par rapport aux moteurs.

Nota importante : Si tout cela vous paraît complexe et pour plus d'informations sur la problématique du référencement de sites réalisés avec des frames, merci de vous reporter à l'article d'avril 2002 de la lettre R&R, intitulé "Comment référencer un site web réalisé avec des frames ou en Flash".

3. Texte des pages

Le texte des pages est géré de façon un peu spécifique.

Sur la page d'accueil "light" :

<http://www.cite-sciences.fr/francais/indexLIGHT.htm>

Pas de problèmes : les liens sont au format texte et il y a assez de contenu textuel en dehors des liens pour être compris par les moteurs. Il s'agit d'une page qui s'apparente à un plan du site, toujours très efficace pour le référencement car proposant de nombreux liens vers les pages internes.

En revanche, dans les pages internes, le contenu est géré par une applet Java, présente à l'adresse http://www.cite-sciences.fr/francais/ala_cite_menu.htm, qui semble afficher le contenu adéquat en fonction de la demande initiale (du clic effectué par le visiteur).

Dans ce cadre, il nous semble que le contenu "informatif" des pages du site ne sera pas indexé par les moteurs, ce qui est un petit peu dommage.

Du coup, ce type de "handicap" doit être compensé, pour obtenir un bon référencement par :

- De bons titres pour les pages "filles"
- Des balises "meta" qui, si leur poids est aujourd'hui moins important, doivent cependant être optimisées si la page ne contient pas beaucoup de contenu textuel.
- Une bonne utilisation de la balise "Noframes" (voir article d'avril 2002) qui ne semble pas gérée sur le site de la Cité des Sciences.

4. Balises Meta

Les balises Meta sont bien gérées sur l'ensemble du site.

Sur la page de redirection, elle ont la forme suivante :

```
<meta name="description" content="La Cité des Sciences à Paris. Visite des expositions scientifiques pour apprendre et pour jouer. La sortie des enfants et des parents. Un musée interactif. Des expositions scientifiques, des colloques, des conférences. Des vidéos et des expérimentations en ligne">
```

```
<meta name="keywords" content="cité des sciences, villette, exposition apprendre, science paris, enfants sorties, Sciences musée, éducation en ligne, jeux interactifs, expériences interactives, cité Science villette, savoir apprendre Paris, colloques et conférences, actualité astronomique, actualité scientifique">
```

La balise meta "description" est cependant un peu longue (ne pas dépasser les 150 à 200 caractères si possible).

La page d'accueil "Light", bizarrement, n'a pas de balise meta. Dommage, car cette page est intéressante et importante pour les moteurs.

Les pages internes du site ont, la plupart du temps, des balises adaptées. Exemple pour la page qui traite des visiteurs handicapés :

```
<meta name="keywords" lang="fr" content="handicap, visiteur handicapé, sourd, muet, aveugle, déficience visuelle, mobilité réduite, fauteuil roulant, handicap mental, Cité des Sciences et de l'Industrie">
```

```
<meta name="Description" content="Accueil des visiteurs handicapés; la Cité;">
```

Pour le planétarium :

```
<meta name="keywords" lang="fr" content="planétarium, astronomie, espace, univers, galaxie, planète, étoile, soleil, terre, Cité des Sciences et de l'Industrie">
```

```
<meta name="Description" content="Spectacle sous une voûte étoilée pour un voyage vers d'autres planètes...">
```

Les balises meta sont donc spécifiques à chaque page, en haut du code HTML. Rien à redire à ce niveau. Le site est bien optimisé, même si certains intitulés de balises pourraient être plus

descriptifs. Mais on peut toujours mieux faire que ce qui a été fait :-)) ("La critique est aisée, l'art est difficile", comme disait Destouches...).

On peut également visualiser dans la plupart des pages une balise "revisit after" :

```
<meta name="revisit-after" content="15 day">
```

Rappelons que cette balise n'est pas utile dans le cadre d'un référencement (on se demande d'ailleurs encore à quoi elle est utile ;-))

5. URLs

Le nom de domaine (cite-sciences.fr) est parfait pour une recherche sur "Cité des Sciences", qui est, a priori, une requête très prisée pour trouver le site. Parfait donc à ce niveau.

Les autres urls pourraient être optimisées pour être plus explicites. Exemple, la page "programme" :

http://www.cite-sciences.fr/francais/ala_cite/cite_pra/agenda/global_fs.htm

La page "handicapés" :

http://www.cite-sciences.fr/francais/ala_cite/cite_pra/accessibilite/global_fs.htm

La page "Planétarium" :

http://www.cite-sciences.fr/francais/ala_cite/spectacl/planetar/global_fs.htm

Dommage, en effet, de ne pas indiquer "planetarium" en toutes lettres plutôt que "planetar", "spectacles" plutôt que "spectacl", etc. Le nom de page "Golbal_fs.htm" n'est pas non plus très parlant.

Mais peut-être existe-t-il des contraintes techniques à l'élaboration et au nommage des pages et des répertoires que nous ne maîtrisons pas à notre niveau. C'est d'ailleurs souvent le cas, selon les technologies utilisées pour créer un site web.

6. Liens

Sur la page "Light" (<http://www.cite-sciences.fr/francais/indexLIGHT.htm>) les liens sont textuels. Le texte des liens représente bien le contenu de la page-cible ("Explora", "Cité des enfants", "Techno cité", etc.), donc tout semble "ok" et bien optimisé.

Un point important : certains liens sont proposés sous la forme :

"Cité des métiers

Cliquez ici"

Avec le lien sur l'expression "Cliquez ici". Du coup, pour les moteurs et Google en particulier, ce sont ces termes qui deviennent "descriptifs" du contenu de la page cible, ce qui est peu parlant. Peut-être aurait-il été plus efficace de mettre le lien sur l'expression "Cité des métiers"...

Sur la page Flash, la plupart des liens ne pourront être suivis par les moteurs (animation Flash, epu appréciée des moteurs) mais cela ne présente qu'un faible caractère de gravité, puisque cette page ne sera, de toute façon, pas indexée par les robots (voir point 1 : redirection).

En revanche, les liens des pages internes du site, gérés par l'application Java, ne devrait pas être suivis par les moteurs (non accessibles dans le code HTML).

7. Conclusion

Le site de la Cité des Sciences est, globalement bien optimisé même si l'applicatif Java d'affichage du contenu textuel des pages internes pourra limiter la prise en compte de mots clés importants par les moteurs. Une attention toute particulière semble avoir été apportée aux balises meta. Les titres des pages et les urls sont peut-être perfectibles pour optimiser le référencement. La problématique du Flash a été bien prise en compte par la page de redirection de départ, qui prend également en considération la problématique du référencement.

Pour information, le site a actuellement 11 200 pages dans Google dont 10 200 en français :
<http://www.google.fr/search?q=-djdhdjhd+site:cite-sciences.fr&ie=ISO-8859-1&hl=fr>

Et 6 594 pages dans Voila.fr :
http://search.ke.voila.fr/S/voila?dt=*&kw=site:cite-sciences.fr

Ainsi que 3 864 documents sur AllTheWeb (<http://www.alltheweb.com/>)

Ce "bon score" est certainement le fruit de la bonne optimisation du site déjà effectuée...