

Le Page Rank: Référencement de Sites

Sommaire

I.	Un peu de culture.....	3
A.	Les moteurs de recherche Internet.....	3
1.	Définition.....	3
2.	Fonctionnement	3
3.	Financement.....	4
B.	Rapide présentation des principaux moteurs de recherche.....	5
1.	Yahoo.....	5
2.	MSN.....	5
3.	Altavista.....	5
4.	Google	5
C.	Comparaison des principaux moteurs de recherche	6
1.	Comparaison d'efficacité.....	6
2.	Comparaison de l'efficacité par critère.....	6
3.	Interprétation	7
II.	Le Page Rank	8
A.	Description.....	8
1.	Définition.....	8
2.	Formule du Page Rank	8
3.	Exemples d'application	10
B.	Le promeneur aléatoire.....	13
C.	Augmenter son Page Rank	15
1.	Obtenir des liens	15
2.	Structurer les liens internes.....	15
III.	En conclusion.....	15
	Annexe : Les critères de recherche de Google.....	16
IV.	Sources :.....	18

I. Un peu de culture...

A. Les moteurs de recherche Internet

1. Définition

Un moteur de recherche est un outil de recherche sur le web constitué de « robots » qui parcourent les sites régulièrement et de façon automatique (sans intervention humaine, ce qui les distingue des annuaires) pour répertorier de nouvelles adresses Internet (URL). Ils suivent les liens hypertextes, reliant les pages les unes aux autres, rencontrés sur chaque page atteinte. Chaque page identifiée est alors indexée dans une base de données, accessible ensuite par les internautes à partir de mots-clés.

Il est également commun d'appeler moteur de recherche :

- des sites web proposant des annuaires de sites web : dans ce cas, ce sont des ressources humaines qui répertorient et classifient des sites web jugés dignes d'intérêt et non des robots d'indexation - on peut citer par exemple Voilà et Yahoo!, etc.
- des logiciels installés sur un ordinateur personnel : ce sont des moteurs dits desktop qui combinent la recherche parmi les fichiers stockés sur le PC et la recherche parmi les sites web - les plus connus et utilisés à l'heure actuelle sont Google Desktop et Copernic Desktop Search.

2. Fonctionnement

Le fonctionnement d'un moteur de recherche se décompose en trois étapes :

- Le Web est continuellement exploré par un robot d'indexation suivant systématiquement tous les hyperliens qu'il trouve et récupérant les ressources jugées intéressantes. Classiquement, l'exploration est lancée à partir d'un annuaire web.
- L'indexation des ressources récupérées consiste à extraire les mots considérés comme significatifs (presque tous) correspondant à chaque ressource. Les mots extraits sont enregistrés dans une base de données organisée comme un gigantesque dictionnaire inverse. La partie requêtes du moteur de recherche peut ainsi rapidement retrouver les correspondances. C'est alors que tous les algorithmes comme le Page Rank, interviennent : un algorithme gardé secret est appliqué pour donner un poids variable aux correspondances, afin de pouvoir présenter les résultats des recherches par ordre de pertinence calculé (qui ne correspond d'ailleurs pas nécessairement à la réalité, puisque les algorithmes ont chacun leur méthode de classification). L'algorithme tient généralement compte du contexte du mot clé (titre, paragraphe, hyperlien...) et de la ressource (ressources liées, popularité du site...)
- La ressource indexée peut ensuite être retournée dans les résultats, suivant les mots clefs recherchés par un utilisateur.

3. Financement

Dans la grande majorité, les moteurs de recherche sont financés grâce à de la publicité ciblée. Classiquement, il s'agit d'afficher sur la page de résultats de recherche des publicités pertinentes suivant la recherche effectuée par l'utilisateur. Cependant, il semblerait que certaines sociétés, comme Google, aient adopté une stratégie plus précise. En effet, au travers de ses différents services, Google index et répertorie et garde en base de données de nombreuses informations sur la façon qu'on ses 'clients' d'utiliser sa gamme de logiciels. Ainsi, par recoupement des données, et surement grâce à un algorithme toujours gardé secret, les publicités présentées sont toujours en adéquation avec les goûts et choix usuels de l'utilisateur.

Il y a deux manières de présenter des publicités au sein d'un moteur de recherche :

- en encart séparé
- en l'intégrant aux résultats de la recherche

Pour le visiteur, l'encart séparé se présente comme une publicité classique. L'intégration aux résultats se fait en revanche au détriment de la pertinence des résultats et peut avoir des retombées négatives sur la qualité perçue du moteur.

B. Rapide présentation des principaux moteurs de recherche

1. Yahoo

Le premier moteur de recherche à avoir vu le jour est Yahoo!, qui figure encore aujourd'hui comme un des premiers parmi les sites les plus visités. En effet, Yahoo! n'est pas réellement un moteur de recherche, mais un annuaire de sites web (et donc pas de pages uniques) classés par sujet. La structure des sites sur Yahoo! est faite de catégories organisées "en arbre": cela permettant aux utilisateurs d'obtenir aisément une subdivision des sites par typologie, en isolant ainsi seulement ceux relatifs à l'argument recherché.

2. MSN

Il fut un temps où MSN, anciennement appelé Microsoft Network, n'avait à l'origine pour unique vocation que de concurrencer America On Line (AOL) et CompuServe aux Etats-Unis. A la suite de l'échec commercial de ce service, MSN s'est vu présenté sous un nouveau jour, cette fois en tant que portail de service. C'est ainsi qu'est né MSN Search, qui plus qu'un moteur de recherche, est un catalogue de liens vers les services de Microsoft comme par exemple Expedia ou Carpoint.

3. Altavista

Mis en ligne à l'adresse Web altavista.digital.com en décembre 1995 par des chercheurs de Digital Equipment Corporation, Altavista fut le premier moteur de recherche à indexer une bonne partie des pages Web existantes et devint immédiatement très populaire. Il fut également le premier moteur de recherche multilingue (la version française fut ouverte le 15 février 2000), ainsi qu'à lancer la recherche d'images, de fichiers audio et de vidéos.

4. Google

Google est une société fondée le 7 septembre 1998 par Larry Page et Sergey Brin, auteurs du moteur de recherche Google. Google s'est donné comme mission d'« organiser l'information à l'échelle mondiale et de la rendre universellement accessible et utile ».

A ce jour, Google aurait indexé plus de 8 milliards de pages Web, 1 milliard d'images et 1 milliard de messages Usenet. En novembre 2005, les créateurs de Google annonçaient avoir multiplié par mille le nombre de pages indexées depuis la création du moteur de recherche, soit un total de 24 milliards de pages.

C. Comparaison des principaux moteurs de recherche

Par soucis de crédibilité, notons que les informations ci-dessous datent d'avril 2005, et sont susceptibles de ne plus être pertinentes aujourd'hui. Les notes présentées ci-dessous sont comptabilisées sur une base de 10.

1. Comparaison d'efficacité

Comparaison d'efficacité				
	Google	Yahoo	msn	
Moyenne générale	7.55	6.80		5.46
Scénarios simples	8.08	7.18		6.00
Scénarios complexes	6.75	6.22		4.64
Recherches marchandes	7.06	7.20		6.08
Recherches de contenu	7.65	6.66		5.27
Recherches d'actualité	7.80	7.00		5.55

Source : Journal du Net, avril 2005

Selon ce comparatif, Google est le moteur qui répond le mieux aux requêtes des internautes, obtenant des notes supérieures à celles des deux autres solutions, dans la majorité des recherches effectuées. Sa moyenne générale est ainsi supérieure de 0,75 point à celle de Yahoo et de 2,09 points à celle de MSN Search, dont les performances sont nettement moins bonnes dans pratiquement tous les cas.

2. Comparaison de l'efficacité par critère

Comparaison d'efficacité par critères				
	Google	Yahoo	msn	
Pertinence des trois premiers résultats	6.62	5.78		3.38
Pertinence des sept derniers résultats	5.60	4.92		3.94
Taux de résultats non redondants	7.40	7.56		5.54
Taux de liens non brisés	8.86	7.46		7.04
Taux de pages non parasites	8.70	8.26		7.06

Source : Journal du Net, avril 2005

Sur la seule pertinence des trois premiers résultats, l'écart se creuse entre les trois moteurs. Google arrive en tête, à 0,84 point de Yahoo et à 3,24 points de MSN Search, qui n'obtient pas la moyenne sur ce critère essentiel. Yahoo se laisse surtout distancer (1 point en dessous) dans

Rapport de VA

le cas de scénarios complexes mais dépasse son concurrent (de 0,4 point) dans le cas de recherches marchandes.

Dès le quatrième résultat présenté, les outils de recherche, qui mettent en avant la taille importante de leur index, fournissent tous des réponses de moins bonne qualité. Ainsi, sur la pertinence des sept derniers résultats, Google obtient une note inférieure d'un point à celle qu'il affiche sur le critère précédent, offrant l'opportunité à ses rivaux de réduire l'écart.

A noter également que Yahoo est le moteur qui fournit les résultats les moins redondants.

Enfin, avec une note de 8,70, Google est également particulièrement bon pour éliminer les pages.

3. Interprétation

Il ressort de ces deux comparatifs que Google, pourtant né plus tard que Yahoo, obtient de bien meilleurs résultats que ses deux principaux rivaux. Sa performance, nettement supérieure lors de recherches 'simples', et donc plus couramment utilisées, combiné à son interface épurée, en font aujourd'hui le moteur de recherche référence pour la plupart des internautes.

Mais quelle est la raison d'une telle démonstration de performances ?

Réponse : le **Page Rank**, algorithme utilisé par Google pour indexer les pages Web. Ce même Page Rank que nous allons tenter d'analyser dans la suite de ce document...

II. Le Page Rank

A. Description

1. Définition

A l'origine, le Page Rank est un critère mathématique permettant de mesurer la popularité d'une page sur le web. Plus il est élevé, plus la page est censée être populaire (et donc intéressante pour l'internaute). Le Page Rank est l'un des éléments pris en compte par Google, mais ce n'est pas le seul. Il en existe sûrement plusieurs dizaines (voir Annexe).

Google affecte ainsi une note de popularité à chaque page Web. Cette note est faite en fonction des liens externes (popularité de liens) pointant vers elle et des liens qu'elle fait vers elle-même (liens internes). Dans ce calcul est intégré un des algorithmes Page Rank pour donner une note stable à chaque page.

Le Page Rank réel d'une page n'est connu que de Google, mais grâce à la Google Toolbar, il est possible d'obtenir une approximation du classement de la page dans la zone Page Rank de la barre ; cette valeur est échelonnée entre 0 et 10 (qui constitue une échelle logarithmique).

Cependant, le Page Rank est utilisé par Google dans tout son processus. Ainsi, un site ayant une ou plusieurs pages au Page Rank élevé aura beaucoup d'avantages :

- Indexation plus fréquente, et donc ses mises à jour seront rapidement disponibles dans Google ; il pourra éventuellement avoir des pages dynamiques indexées ce qui ne serait pas le cas si son Page Rank était plus faible
- Classement en tête de sa catégorie dans l'annuaire de Google ;
- etc.

2. Formule du Page Rank

Mettons nous en situation : imaginons qu'une page A fasse un lien vers une page B ; nous pouvons dès lors conclure que la page A, de par son contenu et sa conception, juge que la page B est suffisamment intéressante et importante dans le contexte pour qu'un visiteur de la page A aille la visiter, en complément où homologie avec cette dernière. Ce simple lien, dit rentrant, de A vers B augmente le Page Rank de B. Ce concept est la base fondamentale de la formule du Page Rank telle qu'elle est utilisée par Google.

Deux idées supplémentaires mais essentielles viennent la compléter :

- Plus le Page Rank de la page A sera haut, plus l'augmentation du Page Rank de la page B sera important en proportion. En d'autres termes, il est bien plus efficace d'avoir un lien depuis la page d'accueil de Google (Page Rank de 10) que depuis une page d'un site lambda avec un Page Rank de plus faible degré.
- L'augmentation de Page Rank de la page B est d'autant plus importante que la page A fait peu de liens sortants (liens pointant vers une autre page). En d'autres termes, si la page A (ou du moins son concepteur) juge qu'il n'y a qu'une page qui mérite un lien, alors il est normal que le Page Rank de la page B augmente plus que dans le cas où de nombreuses pages obtiennent un lien.

Rapport de VA

Penchons nous à présent sur la formule mathématique réelle du Page Rank. Les informations qui suivent sont basées sur un article rédigé par les deux fondateurs de Google, Larry Page et Sergey Brin, même si depuis l'algorithme a dû évoluer : la base reste la même.

- Soient A_1, A_2, \dots, A_n : n pages pointant vers une page B .
- Notons $PR(A_k)$ le Page Rank de la page A_k
- $N(A_k)$ le nombre de liens sortants présents sur la page A_k
- 'd' un facteur compris entre 0 et 1, fixé en général à 0,85. Ce coefficient, un peu mystérieux, sert en fait de facteur d'amortissement.

Alors le Page Rank de la page B se calcule à partir du Page Rank de toutes les pages A_k de la manière suivante :

$$PR(B) = (1-d) + d \times (PR(A_1) / N(A_1) + \dots + PR(A_n) / N(A_n))$$

Cette formule n'est à la base pas très compliquée. Cependant, si elle est simple dans sa conception, elle l'est beaucoup moins dans son application directe et à échelle humaine.

Note : A propose de 'd'

- 'd' est un coefficient d'amortissement choisi arbitrairement entre 0 et 1 (généralement 0.85), permettant de limiter l'importance de l'addition de tous les facteurs $PR(A_n)/N(A_n)$.
- Le facteur $(1-d)$ permet de garantir que la moyenne des Page Rank de l'ensemble des pages du web sera de 1.
- Comme nous le verrons plus tard, la valeur arbitraire prise pour 'd' permet d'accélérer ou non les calculs.

La formule du Page Rank est donc simple parce qu'elle ne dépend que de quelques termes, mais compliquée parce qu'elle réursive : pour calculer le Page Rank d'une page, il faut avoir calculé celui de toutes les pages pointant vers elle.

Mais alors comment commencer ?

Comme il faut bien se donner un point de départ pour commencer la première itération, il suffit de prendre avec des valeurs arbitraires de Page Rank (par exemple 1). Le choix de cette valeur n'a pas d'influence sur le résultat final à condition que toutes les pages commencent avec la même valeur. Une application de la formule permet de calculer un nouveau Page Rank pour chaque page, plus proche de la réalité que la valeur par défaut que nous avons choisie au début.

Ensuite nous recommençons à calculer les Page Rank de toutes les pages en appliquant de nouveau la formule, mais en partant cette fois des valeurs que nous avons calculées précédemment. Après un certain nombre d'itérations, le système converge : les valeurs de Page Rank de chaque page ne bougent plus entre deux itérations.

Normalement, la convergence est obtenue au bout de plusieurs dizaines d'itérations. Cette convergence dépend de deux principaux

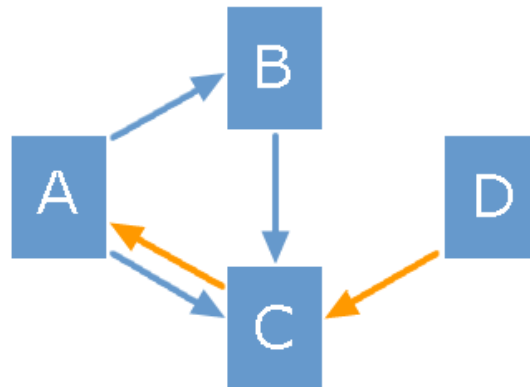
Rapport de VA

- le nombre de pages considérées (important dans la réalité, mais que nous prendrons faible dans nos exemples)
- le facteur 'd' : en effet s'il est choisi trop élevé, le calcul demandera un nombre d'itérations énorme, alors que s'il est trop bas les valeurs ne convergeront pas véritablement, mais finiront par osciller autour de la valeur théorique vraie, un peu à la manière d'un pendule. Avec un facteur d'amortissement de 0.85, il nous faut une quarantaine d'itérations pour affiner le calcul du Page Rank.

3. Exemples d'application

Dans les exemples qui suivent, nous prendrons arbitrairement un coefficient 'd'=0.85, et un Page Rank de 1.0 pour toutes les pages lors de la première itération. Comme nous l'avons vu précédemment, prendre un Page Rank de même valeur pour toutes les pages ne change rien au résultat final, puisque ce dernier converge toujours vers le Page Rank réel des pages, par au dessus ou en dessous.

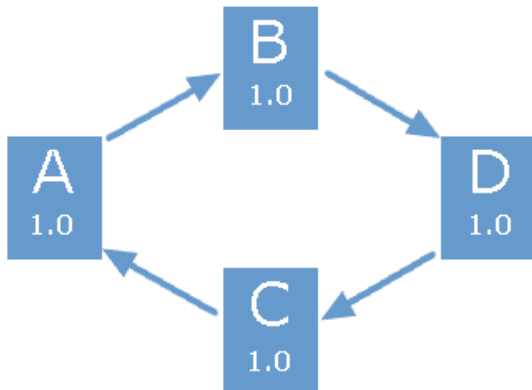
Quatre pages liées :



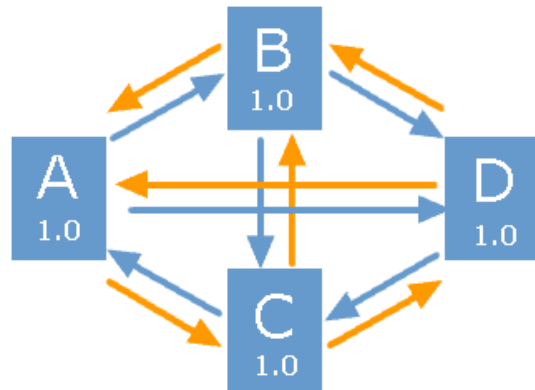
Page	Page Rank
A	1.49
B	0.78
C	1.58
D	0.15

Comme la page D n'a pas de lien entrant son Page Rank est $1-d = 0.15$. De plus, la page C possède le plus haut Page Rank, puisqu'elle profite des 3 liens entrants fournis par les pages A, B et D, n'émettant elle-même qu'un seul lien vers la page A.

Liens circulaires & pages toutes reliées :



Liens circulaires

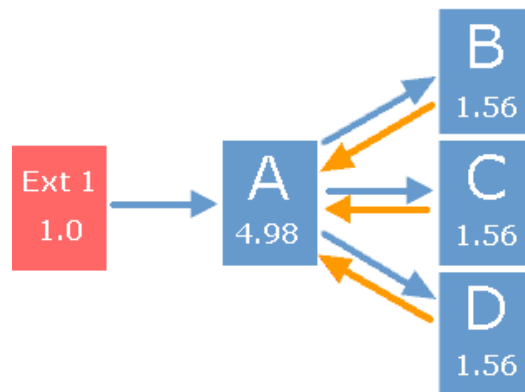


Toutes les pages sont reliées les unes aux autres

Page	Page Rank
A	1.0
B	1.0
C	1.0
D	1.0

Dans ces deux types de configurations, les pages profitent des mêmes gains de Page Rank les unes envers les autres. Elles ont donc toutes un Page Rank de 1.0.

Structure hiérarchique avec un lien entrant :

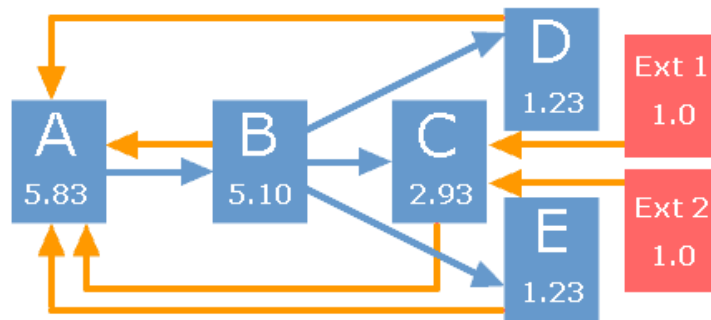


Page	Page Rank
A	4.98
B	1.56
C	1.56
D	1.56

Dans cette configuration nous pouvons constater un gain considérable de Page Rank de la page A. Cela est essentiellement dû au fait que nous considérons ici que la page externe à

Rapport de VA

notre site ne possède aucun lien vers l'extérieur à part nous, ce qui n'arrive en réalité jamais. Cependant, cet exemple illustre parfaitement le fait que moins une page possède de liens sortants, pour les pages pointées par ces mêmes liens profiterons d'une forte augmentation de Page Rank.

Plan de site avec liens entrants sur une page interne :

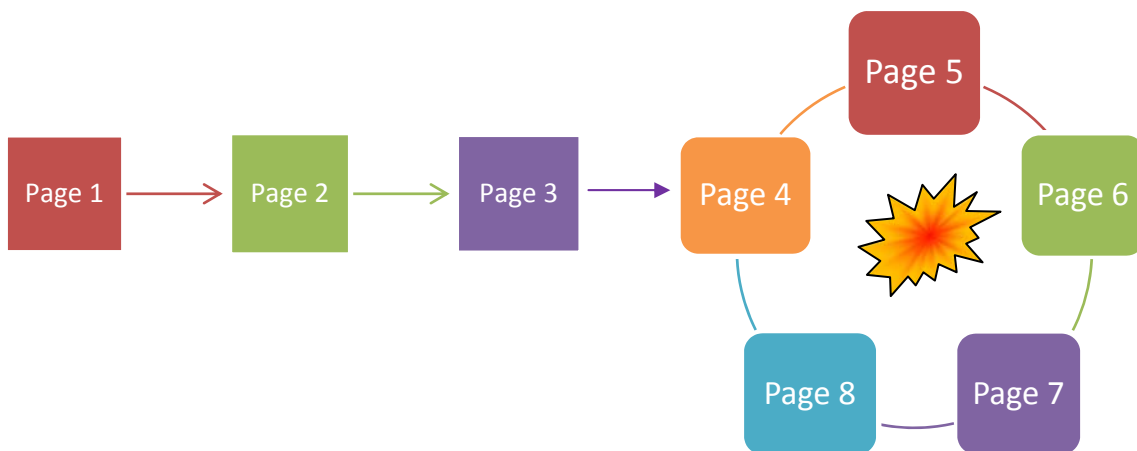
Page	Page Rank
A	5.83
B	5.10
C	2.93
D	1.23
E	1.23

Grâce aux liens externes direct, la page C à un fort gain de Page Rank. Mais il est également intéressant de constater à quel point les pages d'accueil du site (A & B) profitent du Page Rank des pages C, D & E (A profite également de B mais dans une moindre mesure car elle possède des liens sortants vers trois autres pages).

B. Le promeneur aléatoire

Il semblerait que le Page Rank puisse être interprété de différentes façons. L'une d'entre elles (qui était d'ailleurs peut-être la base de la formulation de l'algorithme !), est basée sur la modélisation du surf aléatoire d'un internaute sur le web.

Imaginons que le robot utilisé pour formaliser le Page Rank soit en fait un être humain, en clair, un internaute comme vous et moi. Cet internaute se balade sur la toile en parcourant des pages, cliquant au hasard sur les liens disponibles sur les pages. Il peut ainsi visiter un nombre important de liens, mais également se retrouver cantonné à un groupement de pages reliées les unes aux autres, mais n'ayant pas de lien sur le reste du web. Dans ce cas, il saute aléatoirement sur une autre page web. Il se peut également que, lassé des pages qu'il est en train de visiter, l'internaute décide purement et simplement de charger une nouvelle page, toujours aléatoirement.



- 1- Le promeneur atteint un groupement de pages cyclique. Il saute donc sur une autre page, aléatoirement.



- 2- Le promeneur décide de charger aléatoirement une nouvelle page.



Rapport de VA

Ainsi, le Page Rank d'une page peut être vu comme la probabilité qu'à un instant donné ce surfeur soit précisément sur cette page.

Cette probabilité sera d'autant plus forte que de nombreuses autres pages le ramènent dessus, surtout si ces pages sont elles-aussi importantes : c'est le deuxième terme de la formule.

Le premier terme modélise la probabilité qu'il reste sur la même page, sans suivre de lien sortant. Le facteur d dans la formule peut donc être vu comme la probabilité que le surfeur saute sur une autre page. C'est aussi un facteur indiquant le pourcentage de Page Rank "transmis" aux pages liées.

C. Augmenter son Page Rank

Afin d'augmenter le Page Rank de votre site et donc de vos pages, il y a deux choses à faire :

- Obtenir de liens avec d'autres sites
- Bien structurer les liens internes du site

1. Obtenir des liens

Il s'agit ici d'obtenir le plus grand nombre de liens possibles, de sites externes vers vos pages. Pour ce faire, il y a plusieurs principes à retenir :

- Essayer d'obtenir un grand nombre de liens vers votre page
- Choisir en priorité des pages ayant un bon Page Rank (attention, il s'agit du Page Rank de la page qui pointera vers les vôtres pas du Page Rank du site). Ces mêmes pages doivent être indexées par Google, ne pas être orphelines, et avoir le minimum de liens sortants possible, afin de minimiser l'affaiblissement du Page Rank.

2. Structurer les liens internes

Structurer ses liens internes est aussi, si ce n'est plus important que d'obtenir de liens vers son site. En effet, si les deux solutions peuvent être aussi efficaces l'une que l'autre, il n'en reste pas moins qu'il est très difficile de contrôler les liens entrants. A contrario, la maîtrise des liens internes au site est totale ; c'est pourquoi il est crucial de passer du temps à les ordonner.

Mais pourquoi structurer les liens internes ? Pour la simple raison que les pages d'un même site ont des liens les unes sur les autres, de la même manière qu'une page d'un autre site pointant vers une de vos pages ; l'algorithme du Page Rank ne fait en effet pas la différence entre les liens internes et les liens externes. Cependant, il faut avoir en tête une règle du Page Rank : plus une page a de liens entrants sur elle, plus la part de Page Rank dont elle bénéficiera sera faible.

III. En conclusion...

Le Page Rank reste un algorithme complexe et encore mal connu, parce qu'il a été gardé assez secret par ses concepteurs pour des raisons évidentes d'espionnage industrielle. De nombreuses hypothèses sont émises à son sujet, parmi lesquelles quelques unes, exposées dans le présent document.

Cependant, il serait abusif de croire que seul un Page Rank de marque apportera un grand nombre de visites et une popularité certaine à un site. Ce n'est qu'une partie de la vérité. Le centre névralgique d'un site reste et restera toujours son contenu, qu'il faut essayer de tenir à jour au maximum afin d'inciter les visiteurs à revenir...

Annexe : Les critères de recherche de Google

Voici, présentés ci-dessous, une liste hypothétique des critères utilisés par Google pour classer les pages web. Attention, cette liste est **hypothétique**, et ne peut en aucun cas être considérée comme exhaustive ou sûre. A cet effet, un indicateur de confiance est utilisé pour classer les différents critères.

Les critères de recherche Google	
Critère	Indice de confiance
Page Rank	
Nombre total de liens	Sûr
Nombre de liens contenant le mot-clé dans le texte du lien	Sûr
Rapport entre le nombre de liens et le nombre de liens contenant le mot-clé dans le texte du lien	Hypothèse
PageRank de la page qui fait le lien	Sûr
Nombre de liens dans la page qui fait le lien	Sûr
Nombre de nouveaux liens depuis le dernier calcul du Page Rank	Hypothèse
Balise TITLE	
Présence du mot-clé dans le titre	Sûr
Position du mot-clé dans le titre (distance par rapport au début du texte)	Sûr
Nombre de caractères dans le titre	Hypothèse
Distance entre les mots-clés	Hypothèse
Rapport entre le nombre de mots-clés dans le titre et le nombre total de mots	Hypothèse
Balise META description	
Présence du mot-clé	Hypothèse
Position du mot-clé (distance par rapport au début du texte)	Hypothèse
Rapport entre le nombre de mots-clés et le nombre de mots dans la balise	Hypothèse
Balise META keywords	
Présence du mot-clé	Hypothèse
Position du mot-clé dans le titre (distance par rapport au début du texte)	Hypothèse
Rapport entre le nombre de mots-clés et le nombre de mots dans la balise	Hypothèse
Texte du document	
Position du mot-clé dans le titre (distance par rapport au début du texte)	Hypothèse
Nombre de caractères dans le texte	Hypothèse
Occurrence du mot-clé dans le texte (en dehors de toute balise)	Hypothèse
Indice de densité des mots-clés	Sûr
Distance entre les mots-clés	Sûr
Ordre des mots-clés	Hypothèse
Balise <A>	
Occurrence du mot-clé dans le texte de la balise	Sûr
Balise <H1> ... <H6>	
Occurrence du mot-clé dans le texte de la balise	Sûr
Balise 	
Occurrence du mot-clé dans le texte de la balise	Hypothèse
Présence et imbrication à l'intérieur d'une balise <A>	Hypothèse
Autres balises	
Occurrence du mot-clé dans le texte de la balise	Hypothèse
URL	
Présence du mot-clé dans l'URL	Hypothèse
Position du mot-clé dans l'URL (distance par rapport à la racine du site)	Hypothèse
Nombre de caractères de l'URL	

Rapport de VA

Nombre de niveaux de répertoires	Hypothèse
Fichier	
Date de création	Hypothèse
Taille du fichier en octets (ou Ko)	Hypothèse
Fréquence des mises à jour de la page	Hypothèse
Site	
Taux (ou nombre) de nouvelles pages dans le site depuis le dernier calcul	Hypothèse
Nombre de pages du site	Hypothèse
Page Rank de la page d'accueil	Hypothèse
Pages liantes, liens internes et externes	
Nombre de domaines différents des pages liantes	Hypothèse
Densité des mots-clés des pages liantes	Hypothèse
Présence et position des mots-clés dans les pages liantes (dans le titre, ou le texte)	Hypothèse
Présence des mots-clés dans les pages liées par les pages liantes (pages similaires)	Hypothèse
Densité des mots-clés dans les pages liées par les pages liantes (pages similaires)	Hypothèse
Nombre de liens internes (même url - même répertoire)	Hypothèse
Nombre de liens intermédiaires (même nom de domaine mais autres répertoires)	Hypothèse
Nombre de liens externes (autres noms de domaine)	Hypothèse
Rapport entre le nombre de liens externes et le nombre total de liens	Hypothèse
Nombre de liens externes contenant le mot-clé dans le texte des liens	Hypothèse
Rapport entre le nombre de liens externes et le nombre de liens externes contenant le mot-clé dans le texte du lien	Hypothèse

IV. Sources :

<http://www.wikipedia.org>

<http://www.lesmoteursderecherche.com>

<http://www.webrankinfo.com>

<http://www.journaldunet.com>