

Quelques statistiques sur le Pagerank & les SERP

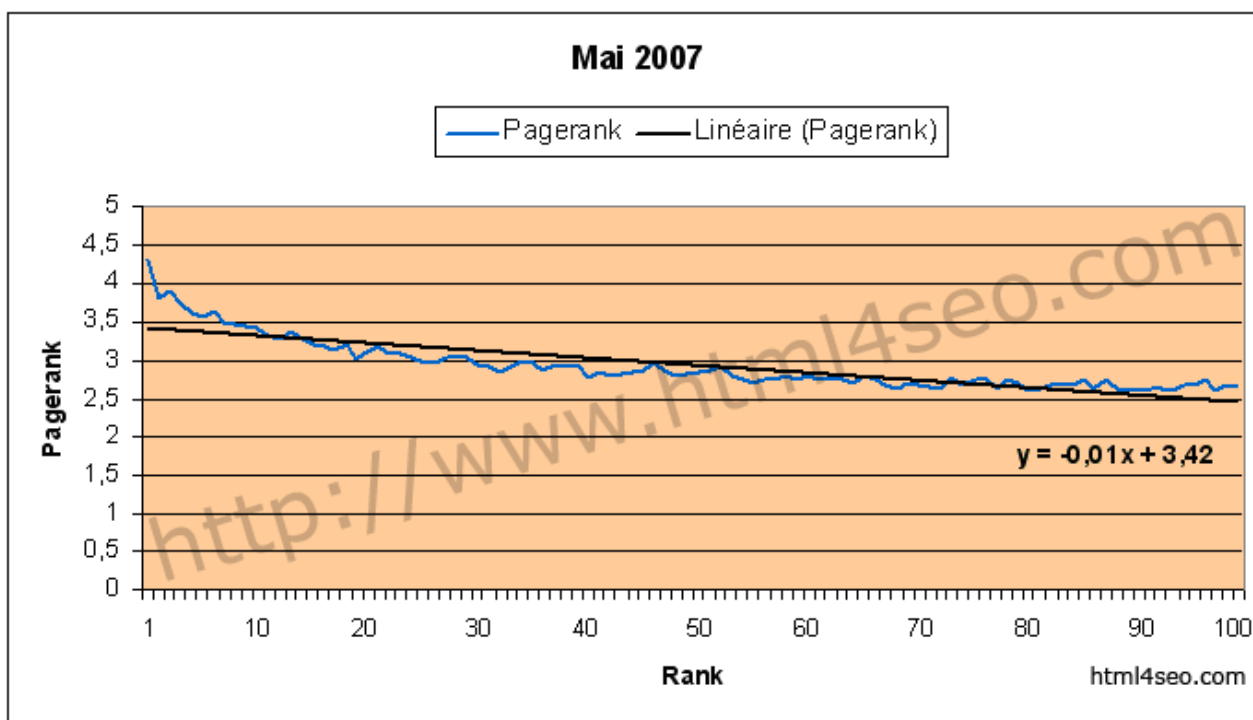
Le Pagerank traqué

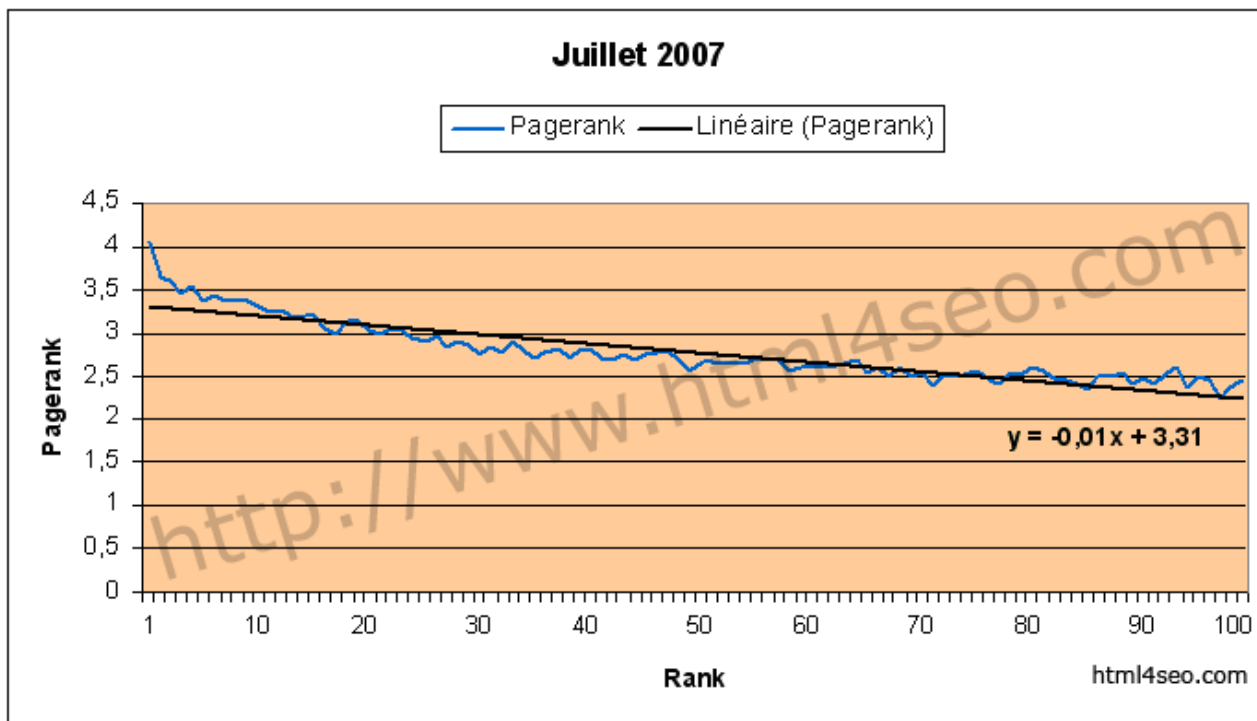
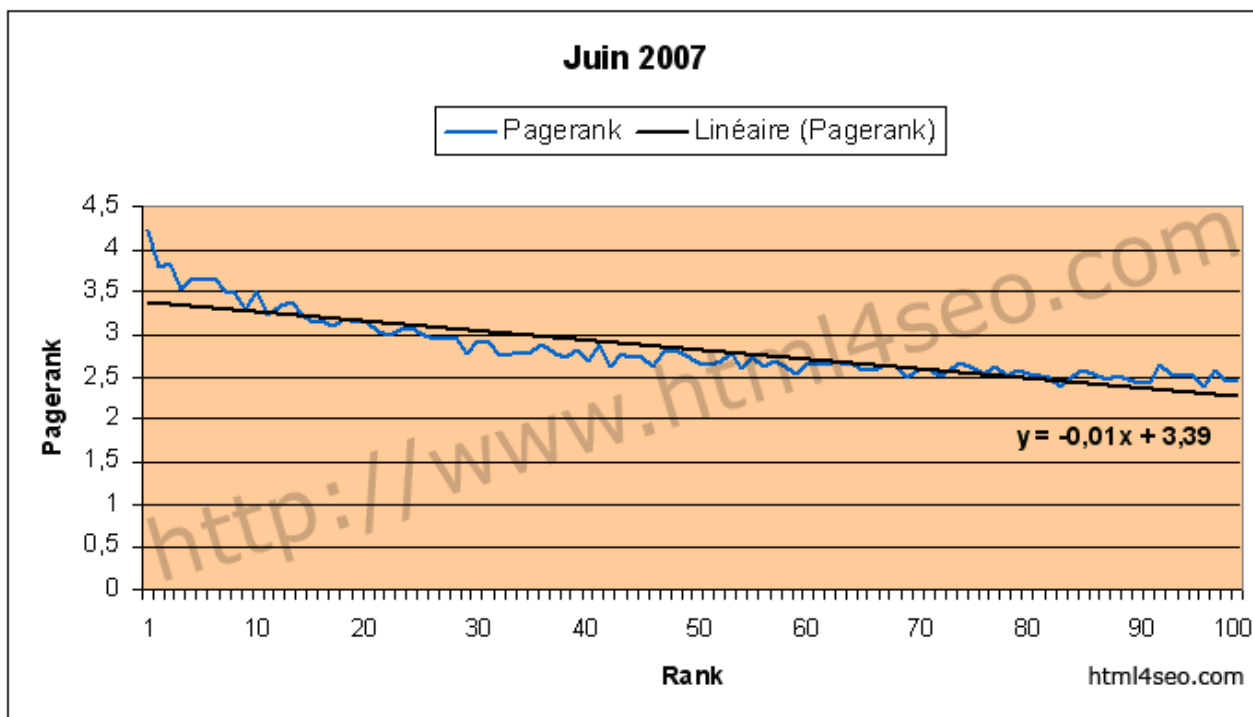
Depuis plusieurs mois l'équipe HTML4SEO (<http://www.html4seo.com>) effectue des milliers d'audits référencement des SERP Google. Dans chacun de ces audits est organisée la collecte du Pagerank, du nombre de backlinks et du contenu HTML pour les 100 premiers des SERP Google. Sur la base de ces résultats, l'équipe HTML4SEO vous propose une petite étude statistique (sans prétention) qui devrait aboutir à quelques conclusions appuyées sur des données concrètes. Nous tâcherons de limiter au strict minimum les données techniques.

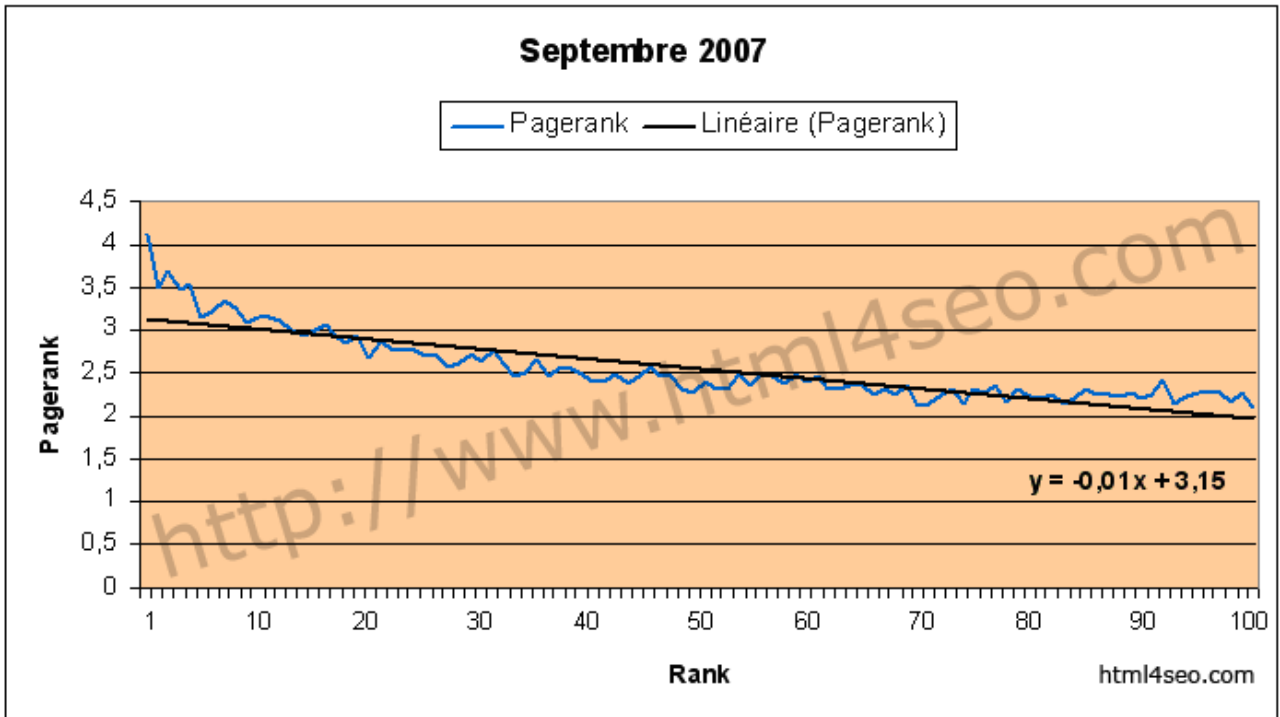
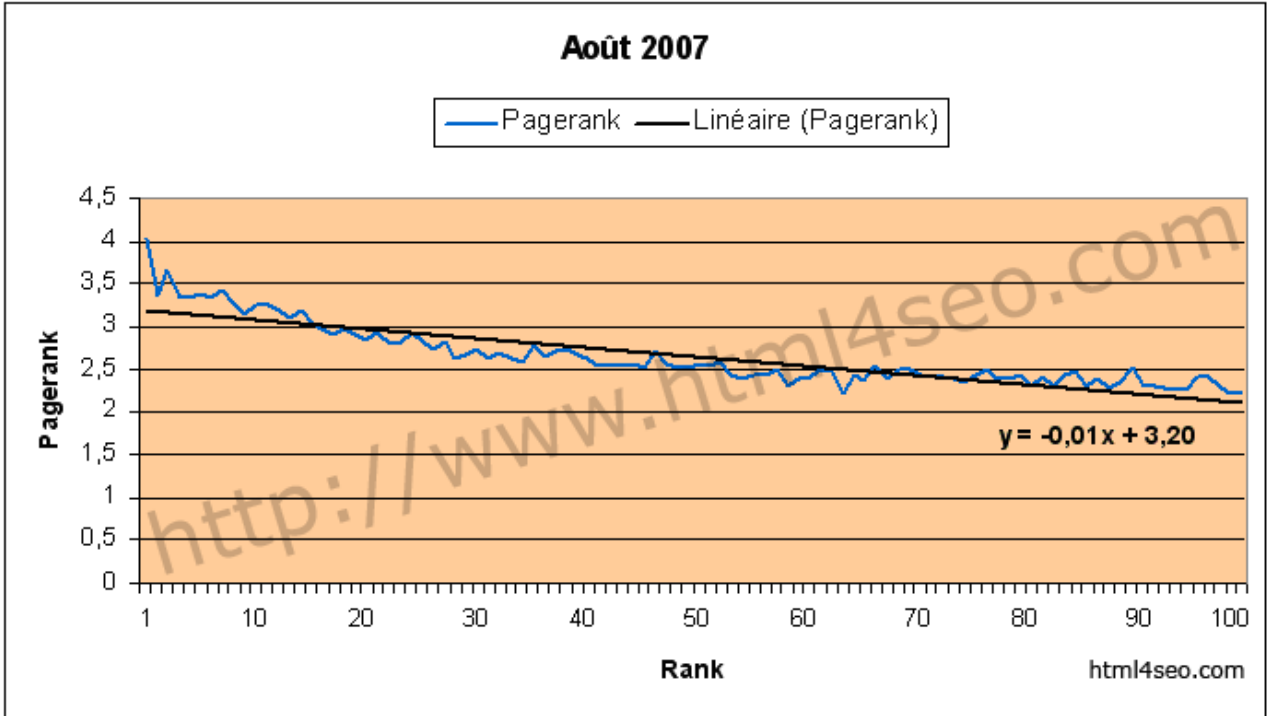
L'étude se restreint au Pagerank, bien que nous aurions pu l'étendre aux backlinks et au contenu HTML des pages recensées dans les SERP Google. Prochainement, nous publierons d'autres études sur ces 2 autres critères majeurs pour le référencement.

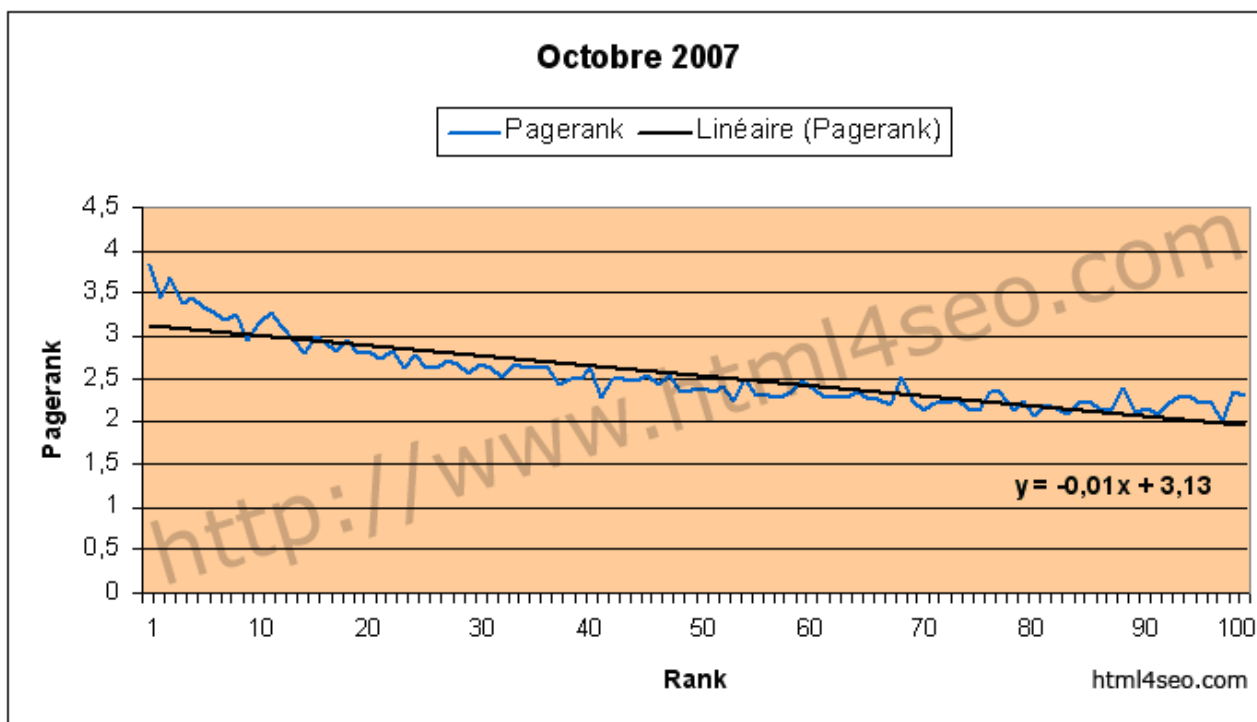
Evolution du Pagerank dans les SERP depuis 6 mois

Graphiques Pagerank









Corrélation Rank/Pagerank

La mesure de corrélation ρ_{XY} entre 2 variables X et Y la plus souvent utilisée est le coefficient de corrélation de Bravais-Pearson souvent appelé coefficient de corrélation

lien [http://fr.wikipedia.org/wiki/Corrélation_\(mathématiques\)](http://fr.wikipedia.org/wiki/Corrélation_(mathématiques))

Ce coefficient de corrélation ρ_{XY} varie entre -1 et 1 et s'interprète comme suit.

Si $\rho_{XY} = 1$ (resp. -1) alors X et Y sont dites parfaitement corrélées positivement (resp. négativement). Connaissant la valeur prise par X, on peut déduire la valeur prise par Y. Les points (X,Y) sont alignés sur une droite de pente positive (resp. négative).

Si $\rho_{XY} = 0$ alors X et Y sont indépendantes ou dépendantes mais non linéairement (la relation qui lie X et Y est plus complexe).

Dans le cas qui nous intéresse X et Y sont respectivement le Rank (position dans les SERP) et le Pagerank. Ci-dessous les données relevées sur les 1000 premiers audits (avec des SERP de plus de 100 résultats) réalisés chaque mois de mai à octobre 2007.

2007	Coef. de corrélation
mai	-0,85
juin	-0,88
juillet	-0,90
août	-0,88
septembre	-0,86
octobre	-0,88
Moyenne	-0,88
Ecart-type	0,02

Commentaires :

1^{ère} observation, le coefficient de corrélation entre le Rank et le Pagerank est proche de 1 en valeur absolue, donc très fort. Le Rank et le Pagerank sont donc fortement corrélés. Un bon pagerank a donc une incidence incontestable sur le positionnement.

2^{ème} observation, le coefficient de corrélation entre le Rank et le Pagerank est très stable sur les 6 derniers mois, autour de -0.88 avec un écart-type à 2%. Le poids qu'occupe le Pagerank dans l'algorithme de Google n'a donc probablement pas varié. Et le Rank est dépendant du Pagerank.

Ceux qui disaient « le pagerank est mort ! » ou « le pagerank, c'est juste de la communication » ont manqué la cible. Le pagerank n'est ni mort, ni réduit à de la communication, il a part importante dans l'algorithme de Google.

Interpolation linéaire Rank/Pagerank

Voyons maintenant les droites de régression linéaire

lien http://fr.wikipedia.org/wiki/Régression_linéaire

	Régression linéaire	Coefficient directeur A	Ordonnée à l'origine B
mai	$y = -0,01x + 3,42$	-0,01	3,42
juin	$y = -0,01x + 3,39$	-0,01	3,39
juillet	$y = -0,01x + 3,31$	-0,01	3,31
août	$y = -0,01x + 3,20$	-0,01	3,20
septembre	$y = -0,01x + 3,15$	-0,01	3,15
octobre	$y = -0,01x + 3,13$	-0,01	3,13

Les valeurs sont données au 1/100^{ème} près.

Commentaires :

1^{ère} observation, le coefficient directeur A

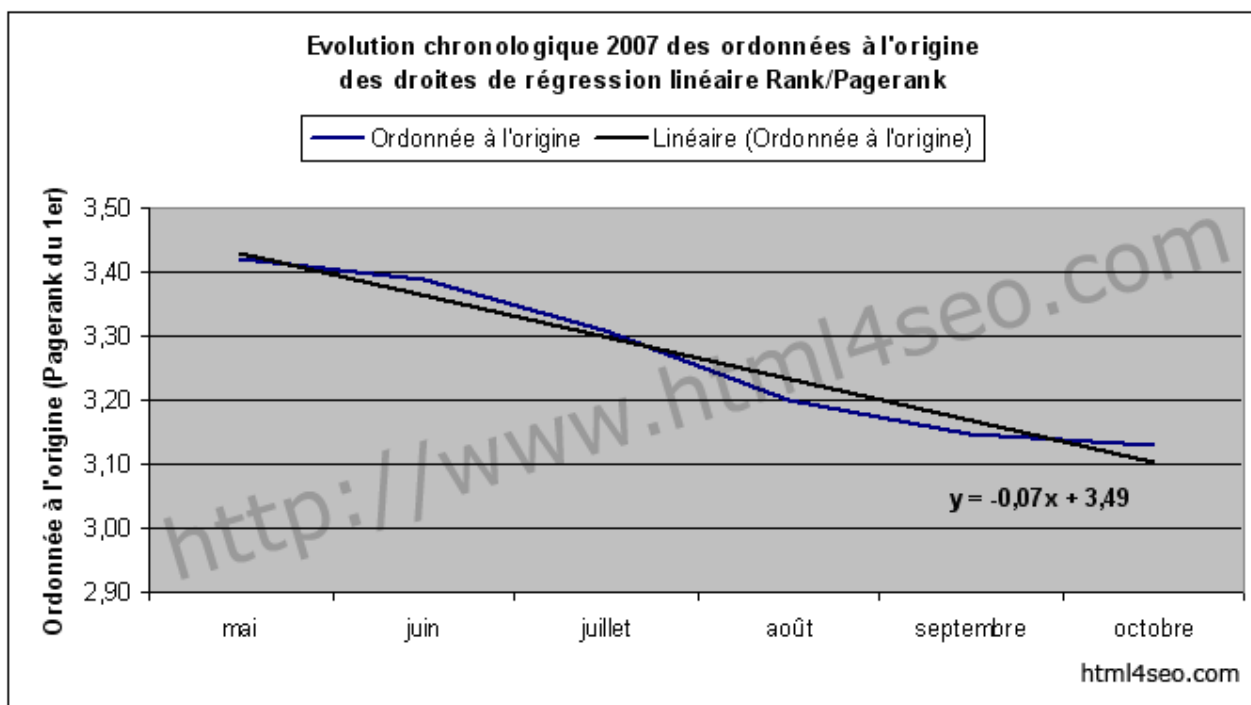
Lien http://fr.wikipedia.org/wiki/Fonction_affine

est constant au 1/100^{ème} près avec une perte de 1 point de pagerank entre la 1^{ère} et la 100^{ème} position qui ne change pas au fil des mois.

2^{ème} observation, l'ordonnée à l'origine

Lien http://fr.wikipedia.org/wiki/Fonction_affine

(interpolation linéaire du pagerank du 1^{er}) décroît. Autrement dit, il semble que les exigences en matière de Pagerank pour parvenir en tête des SERP diminuent mois après mois.



L'étude de la distribution chronologique (de mai à octobre 2007) des ordonnées à l'origine des droites de régression nous donne un coefficient de corrélation de -0.98, soit une corrélation très forte (dépendance quasi-linéaire) entre ces 2 variables. La droite de régression linéaire a pour équation $y = -0.07x + 3.49$, ce qui réduit de 1 point de pagerank tous les 14 mois les exigences en matière de pagerank pour parvenir en tête des SERP.

Ces résultats sont à prendre avec des précautions. Ces données sont statistiques, elles ne donnent donc que des tendances générales. Elles sont néanmoins intéressantes.

Quelle peut-être l'origine de cette perte de pagerank dans les SERP ?

Le web croît et donc les pages web pertinentes à faible pagerank sont plus nombreuses. Dans la quantité de pages web possibles, il y en a pour se distinguer en qualité malgré leur faiblesse en Pagerank/Backlinks. Ceci pose la question de la définition au sens de Google de la pertinence sémantique d'une page web ?

Ou alors, la chasse aux backlinks suspects a commencé il y a quelques mois. Les backlinks considérés fiables étant de moins en moins nombreux, ce sont des pages web mieux sémantiquement (intérieurement : code HTML et extérieurement : Backlinks) ciblées qui passent devant. Et cette évolution du Pagerank des SERP est possible car le Pagerank est une mesure globale, non sectoriellement ciblée.

Ou alors, le réétalonnage du Pagerank (hypothèse émise par certains) a commencé il y a plusieurs mois déjà. Et la mise à jour récente du Pagerank visible viendrait confirmer cette hypothèse. A noter qu'un réétalonnage devrait impliquer une perte globale (quasi-uniforme) de Pagerank.

Il y a encore probablement d'autres hypothèses ...

Le Pagerank du mois de novembre 2007

Apparemment les modifications récentes de Pagerank sont plus importantes qu'aux précédentes mises à jour. Certains ont gagné jusqu'à 3 points de Pagerank alors que d'autres en ont perdus jusqu'à 3. De plus, les modifications de Pagerank ne semblent pas être à sens unique. Etant donné la forte corrélation Rank/Pagerank (voir plus haut), il y a lieu de s'inquiéter pour ceux qui en ont perdu. Car cette perte de Pagerank risque fort de s'accompagner d'un affaiblissement des positions dans les SERP Google et donc d'une perte de trafic.

Nous tâcherons de faire un point le mois prochain sur l'évolution de ces indicateurs.

HTML4SEO Team, Novembre 2007

<http://www.html4seo.com>