

La demande de brevet "Information Retrieval Based on Historical Data" de Google

[Retour au sommaire de la lettre](#)

Google a déposé fin mars aux Etats-Unis une demande de brevet qui suscite de nombreux débats dans le monde de la recherche et du référencement. Compte tenu de la date de parution de cette information, entre le 31 mars et le 1^{er} avril 2005, certains ont tout d'abord cru à un poisson d'avril ! D'autres ont également envisagé qu'il puisse s'agir d'un brevet "leurre" visant à tromper les concurrents ou ils y ont encore vu la preuve écrite de l'existence de la fameuse "sandbox", une théorie selon laquelle Google mettrait en quarantaine dans un "bac à sable" les sites récents lorsque l'apparition de liens pointant vers ces sites serait trop rapide pour être "naturelle" et risquerait d'améliorer injustement leur positionnement. Pour beaucoup, Google souhaite en outre véritablement "cadenasser" sa technologie avec ce brevet afin qu'elle ne puisse être utilisée par ses concurrents. Bien qu'il ne soit en rien certain que toutes les technologies qui sont exposées dans cette demande soient actuellement utilisées par Google, ce document apporte plusieurs éclairages sur le fonctionnement de ce moteur, d'ordinaire extrêmement "secret" quant à sa technologie. Nous vous proposons ici un récapitulatif du contenu de la demande et des débats qu'elle suscite.

Le résumé de cette demande de brevet, déposée auprès de l'*United States Patent and Trademark Office* (USPTO), indique qu'elle porte sur l'utilisation des **données historiques** pour déterminer le positionnement d'un document : "Un système identifie un document et obtient une ou plusieurs sortes de données historiques liées à ce document. Ce système peut ensuite attribuer une note [un score] à ce document en fonction, au moins partiellement, de cette ou de ces donnée(s) historique(s)" (voir [Cette demande de brevet couvre un très large éventail de techniques de classement \(*ranking*\) et de méthodes permettant d'évaluer la pertinence des documents. Selon la méthode décrite, trois principaux critères sont recherchés et évalués afin de déterminer la pertinence d'un document : son **âge**, les **changements** qui ont été effectués et les **tendances** \(pics de popularité...\). Ces critères sont à la fois appliqués aux documents eux-mêmes et aux liens pointant vers ces documents.](http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&co1=AND&d=PG01&s1=20050071741&OS=20050071741&RS=20050071741>qui).</p></div><div data-bbox=)

Notez bien qu'il ne s'agit pour l'instant que d'une demande qui pourrait éventuellement se conclure par la délivrance d'un brevet dans un an environ.

Principaux points à retenir

Le premier point à noter est que les méthodes présentées dans ce document permettent d'évaluer la pertinence de tous les documents numériques (qu'il s'agisse de mails, de sites Web, de fichiers multimedia, de blogs ou encore de publicités...) et pas seulement des pages Web. Elles pourraient s'appliquer donc non seulement au réseau Internet, mais également à un intranet ou à un outil de "Desktop Search"...

La demande de brevet fait état de plusieurs méthodes :

- Pour définir la date de création / d'enregistrement d'un document

La date de création d'un document est une donnée utilisée par Google pour déterminer la pertinence et la qualité d'un document. Elle est définie comme la date de l'enregistrement d'un site ou la date à laquelle le spider de Google a pour la première fois crawlé ce document.

- Pour définir la fréquence de rafraîchissement des contenus et leurs changements

Google utilise également des données sur la fréquence de rafraîchissement des contenus pour déterminer la pertinence des documents. Les données sur les changements effectués peuvent être utilisées pour déterminer la "fraîcheur" d'un document ou son caractère "statique".

- Pour la détection du spam et l'application de pénalités

Plusieurs systèmes de détection et de prévention du spam sont décrits dans cette demande de brevet. Ces systèmes comprennent la surveillance des sites dont le classement progresse "trop rapidement", la surveillance des données qui ont été "blacklistées" par le moteur (adresses IP, serveurs, etc.)...

Cette demande de brevet valide-t-elle l'effet "sandbox" (bac à sable) ?

Le célèbre algorithme "PageRank" de Google assigne à chaque page un score reflétant son importance en fonction du nombre et de la qualité des liens pointant vers cette page. Pour contrer les abus (en créant notamment des liens "fictifs" pointant vers une page cible pour augmenter son positionnement), de nombreux référenceurs pensent que Google a mis en place un système de pénalité à l'encontre des sites suspectés de tenter de gonfler artificiellement leur PageRank en faisant croître trop rapidement le nombre de liens pointant vers eux. C'est ce que l'on appelle l'effet "sandbox", sorte de "bac à sable" où seraient mis en quarantaine les sites soupçonnés de fraude.

Rien n'indique clairement dans la demande de brevet l'existence de cette "sandbox" mais l'un des articles offre une piste quant à l'existence d'une possible pénalité infligée aux nouveaux sites :

"[0039] Prenons, par exemple un document "né" hier et référencé par 10 backlinks. Ce document sera, peut être, plus valorisé par le moteur de recherche qu'un document âgé de 10 ans référencé par 100 backlinks, et ce pour la simple raison que le taux de croissance de ses backlinks est considérablement supérieur au taux de croissance du document plus ancien. Toutefois, même si une croissance brusque du nombre de backlinks peut entraîner une valorisation du document par le moteur de recherche, cette croissance rapide peut également être le signe d'une tentative de spam contre ce même moteur de recherche. C'est la raison pour laquelle le moteur de recherche peut, au contraire, diminuer la valorisation d'un document avec pour but de se protéger contre le spamming".

Par ailleurs, concernant la prise en compte des liens pointant vers un document pour le calcul de sa pertinence, le composant 59 de la demande indique qu'une pénalité peut être infligée à un document lorsque les liens pointant vers lui ont une durée de vie courte et un bonus accordé si ces liens ont une longue durée de vie. Ce critère, qui considère que plus un lien est ancien et plus il a de valeur, pourrait également expliquer en partie le phénomène du "sandboxing". Affaire à suivre... ;-)

Tableau récapitulatif des données pouvant être analysées pour déterminer la pertinence

Les critères que l'on estime déjà pris en compte par l'algorithme de Google pour déterminer la pertinence d'un document figurent en gras

Éléments contrôlés (composants du brevet)	Principaux critères d'évaluation utilisés
<i>Documents / Contenus</i> (2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10 / 11 / 12 / 13 / 14 / 19 / 20 / 21 / 50 / 51 / 52 / 53)	- Date de création ou d'enregistrement. - Fréquence de changement et de rafraîchissement des contenus. - Documents en voie de "péremption" (<i>staleness</i>). - Thématique des documents.
<i>Comportement</i> (15 / 16 / 36 / 37) <i>et données utilisateurs</i> (45 / 46 / 47)	- Taux de conversion en clics (nombre moyen de clics sur le lien dans la page de résultats). - Temps passé sur un document. - Données provenant des utilisateurs : favoris, cache et fichiers temporaires ; historique des visites de certains documents.
<i>Requêtes</i> (17 / 18)	- Documents associés à certains termes de recherche. - Evolution dans le temps du sens des requêtes.
<i>Liens</i> (22 / 23 / 24 / 25 / 26 / 27 / 28 / 29 / 49 / 54 / 55 / 56 / 57 / 58 / 59 / 60 / 61 / 62 / 63)	- Comportement des liens (surveillance de la croissance du nombre de liens pointant vers une page). - Fraîcheur des liens (pondération des liens en fonction de leur "fraîcheur"). - Analyse des rapports entre les liens qui n'appartiennent pas à la même thématique que les sites pointés.

Eléments contrôlés (composants du brevet)	Principaux critères d'évaluation utilisés
<i>Ancre</i> (contenu textuel des liens) (30 / 31 / 32 / 33 / 48)	- Degré de "fraîcheur" des ancrés. - Évolution comparée des contenus et des ancrés des liens.
<i>Trafic</i> (34 / 35)	- Trafic associé aux documents.
<i>Domaine</i> (38 /39 /40)	- Contacts des responsables et administrateurs. - URL (longueur...), hébergement & société.
<i>Ranking</i> (41 / 42 / 43 / 44)	- Historique des classements préalables (<i>prior ranking</i>) et évolution de ces derniers.

Sources : WebProNews (<http://www.webpronews.com/insiderreports/searchinsider/wpn-49-20050408GoogleUnitedGooglePatentExamined.html>) ; Abondance

Analyse détaillée des 63 composants du brevet

Données historiques

Composant 1 : Google peut utiliser "une ou plusieurs données historiques" sur un document pour évaluer la pertinence de ce document. A partir des données historiques recueillies, le moteur génère un indice de pertinence, un score, pour chaque document.

Dates de création ou d'enregistrement des documents

Composant 2 : La date de création ou la date d'enregistrement d'un document (*inception date*) est l'une des principales données utilisées pour générer un indice de pertinence. La pertinence d'un document est au moins partiellement définie en fonction de cette date de création.

Composant 3 : Lorsqu'un document est composé de plusieurs éléments, Google peut déterminer l'ancienneté de chacun des composants et calculer une "moyenne d'âge" pour les différents éléments du site. La différence entre l'ancienneté d'un document spécifique et l'ancienneté moyenne de tous les documents présents sur un site sera utilisée pour définir l'indice de pertinence et le classement.

Composant 4 : La pertinence d'un document est également fonction du temps qui s'est écoulé depuis sa création, c'est-à-dire de son âge.

Composant 5 : La méthode pour déterminer l'âge d'un site peut reposer sur plusieurs critères : la date de "découverte" du site par Google (premier *crawling* par son *spider*), la date à laquelle Google a répertorié un lien pointant vers ce site ; ou encore la date à laquelle le site a atteint un "nombre prédéterminé de pages".

Note : Google privilégie-t-il dans son classement les sites qui ont atteint un certain seuil en termes de nombres de pages ? Autrement dit, les "gros sites" au détriment des sources d'informations moins pourvues en contenu ?

Fréquence des changements et rafraîchissement des contenus

Composant 6 : La fréquence de rafraîchissement des contenus est une donnée utilisée, au moins partiellement, pour définir la pertinence d'un document.

Composant 7 : La fréquence de renouvellement des contenus sera déterminée par le temps moyen écoulé entre les changements, par le nombre de changements effectués dans un intervalle de temps précis ainsi que par la comparaison entre les "taux "de changement effectués sur une période récente et sur une période plus ancienne.

Composant 8 : Le calcul de la pertinence reposera également sur le degré de changements effectués dans un site pendant une période donnée (pages nouvelles, changements, etc...).

Composant 9 : Le calcul de la pertinence à partir des changements sera déterminé en fonction du nombre de pages nouvelles dans une période donnée, de la proportion de pages nouvelles comparée aux pages anciennes et du pourcentage du contenu d'un document qui a changé pendant une période donnée.

Composant 10 : Le calcul de la pertinence à partir des changements sera fonction de l'importance perceptible des changements effectués. La pertinence sera aussi liée au poids des pages concernées par des changements [L'importance d'un document est principalement fonction du nombre de liens (internes ou externes) qui pointent vers lui]. En d'autres termes, des changements sur des pages importantes auront plus de valeur que des changements sur des pages secondaires.

Composant 11 : Etablir la pertinence "d'une pluralité de documents" – i.e. de nombreuses pages dans un site Web donné – implique de déterminer la dernière date de changement pour chaque page, de déterminer la date de changement moyenne et d'établir un indice de pertinence en fonction de la différence entre les deux dates.

Note : cette volonté de prendre en compte l'historique des changements du contenu d'une page est nouvelle, à notre sentiment. Il semble s'agir ici d'une volonté claire de Google de lutter contre le spam... Mais la puissance de calcul nécessaire pour effectuer tous ces travaux au quotidien nous semble phénoménale...

Nombre de changements sur une période donnée

Composant 12 : Le calcul de la pertinence par Google peut aussi inclure une mesure du volume de contenu qui change sur un site pour une période donnée.

Composant 13 : Le volume de changement de contenu sera déterminé par la proportion de pages nouvelles comparée au nombre total de pages sur le site et par le pourcentage de changements dans le contenu pour une période donnée.

Composant 14 : La pertinence des changements sur cette période sera calculée en "mesurant des portions différentes de contenu selon leur importance perceptible".

Taux de clics

Composant 15 : Les données historiques pourront inclure des informations sur la fréquence à laquelle le document est sélectionné lorsque celui-ci est inclus dans un ensemble de résultats de recherches. Cette évaluation fait, au moins partiellement, partie du calcul de la pertinence, les sites ayant le plus grand nombre de clics étant récompensés.

Composant 16 : Google peut privilégier un document lorsque celui-ci est sélectionné plus souvent que les autres documents dans l'ensemble des résultats de recherche pour une période donnée.

Note : Google récompense-t-il les sites enregistrant les meilleurs taux de clics ? Les composants 15 et 16 du brevet tendent à confirmer l'idée selon laquelle le moteur effectue un "monitoring" du nombre de clics par document pour privilégier dans ses résultats les sites qui enregistrent les taux de clics les plus importants. On sait que Google effectue, de façon temporaire, ce type de test dans ses pages de résultats. Ce système, baptisé "indice de clic", avait été popularisé par la société DirectHit, à une certaine époque, notamment sur le moteur Inktomi. Ce n'est donc pas une notion nouvelle...

Documents associés à certains termes de recherche

Composant 17 : Il est possible de déterminer à partir des données historiques les mots-clés qui sont le plus fréquemment utilisés dans les interrogations de recherche et de vérifier si un document correspond à ces termes d'interrogation spécifiques afin d'évaluer sa pertinence pour une requête donnée.

Evolution dans le temps du sens des requêtes

Composant 18 : Selon le brevet, Google calcule si les informations relatives aux interrogations de recherche restent approximativement identiques ou bien si elles évoluent au fil du temps. Le calcul de la pertinence des documents est en partie effectué grâce à ce critère. Par exemple, avant le 11 septembre 2001, les mots clefs «9-11» n'étaient pas été lié au terrorisme alors que c'est aujourd'hui le cas (cf <http://www.socengine.com/seo/guide/google-historical-data-patent.html>).

Documents en voie de "péremption" (*staleness*)

Composant 19 : Le "degré de péremption" des documents peut être évalué pour déterminer la pertinence.

Composant 20 : Google peut aussi déterminer si les documents en voie de péremption sont préférables pour certains types d'interrogations (lorsque le document recherché est, par définition, contraire à la notion de "fraîcheur" ou lorsqu'une réponse unique, spécifique est suffisante).

Composant 21 : Pour évaluer l'intérêt des documents "peu frais/périmés", le moteur peut déterminer à quelle fréquence ces documents "anciens" ont été sélectionnés en concurrence avec des documents plus récents au fil du temps.

Comportement des liens

Composant 22 : Le calcul de la pertinence lié aux données historiques implique d'analyser le "comportement des liens au fil du temps".

Composant 23 : L'apparition et la disparition des liens sont un critère pour l'évaluation de la pertinence.

Composant 24 : L'apparition et la disparition d'un ou plusieurs liens est datée par Google et utilisée dans son calcul de la pertinence.

Composant 25 : La dimension temporelle (basée sur le temps) des groupes de liens est évaluée, avec un contrôle de leurs apparitions/disparitions. Cela permet de déterminer combien de liens associés au document apparaissent ou disparaissent pendant une période donnée et si il y a une tendance à l'augmentation ou à la diminution des liens.

Note : *Il semblerait que cette notion concrétise l'effet "Sandbox"...*

Fraîcheur des liens

Composant 26 : L'évaluation de la pertinence implique de déterminer la fraîcheur des liens associés au document et de leur allouer plus ou moins de poids en fonction de leur "fraîcheur".

Composant 27 : La "fraîcheur" d'un lien est calculée d'après au moins la date d'apparition de ce lien, la date d'un quelconque changement dans l'ancre de ce lien hypertexte (le texte du lien), la date de la page et du site d'où provient le lien, et la date des liens menant à cette page de liens.

Composant 28 : Le poids d'un lien prend aussi en compte le niveau de confiance d'un document contenant un lien, le niveau d'autorité du document où se trouve le lien, ainsi que la fraîcheur du document contenant le lien.

Composant 29 : L'évaluation de la pertinence prend également en compte la répartition par âge des liens renvoyant vers un document. Google prendra en compte l'âge des liens qui pointent vers votre page, ainsi que les périodes précises durant lesquelles ces liens ont été mis en place, par exemple s'agit-il de nouveaux liens ou de liens qui se sont ajoutés régulièrement depuis longtemps ?...

Evolution des ancrés des liens hypertexte

Composant 30 : Pour évaluer la pertinence, les changements survenus au fil du temps dans les ancrés (partie textuelle) des liens hypertextes sont en partie pris en compte.

Évolution comparée des contenus et des ancrés des liens

Composant 31 : Les différences entre les évolutions des contenus des documents et les ancrés associées aux liens pointant vers ces documents sont également mesurées.

Note : *L'analyse comparée des ancrés et des liens permet-elle à Google de détecter les sites "spammeurs" utilisant de "fausses informations" pour attirer les visiteurs ? Par exemple, un texte de lien n'ayant aucun rapport avec le contenu de la page de destination... Ou alors, trop d'ancrés identiques (même contenu textuel) apparaissant en peu de temps sur différents sites, toutes ces ancrés pointant vers le même site peut paraître comme un phénomène suspect...*

Degré de "fraîcheur" des ancrés

Composant 32 : Pour déterminer la pertinence, la fraîcheur des ancrés (textes des liens hypertextes) est l'un des critères utilisables, au moins en partie.

Composant 33 : La fraîcheur de l'ancre est calculée en fonction de la "date d'apparition" de l'ancre, de la "date de changement" de cette ancre, ainsi que par la date de changement et d'apparition de la page où se trouve le lien.

Caractéristiques des documents en termes de trafics

Composant 34 : Les caractéristiques du trafic lié peuvent être pris en considération.

Composant 35 : Les caractéristiques du trafic lié à un document sont analysées afin d'identifier les changements au fil du temps.

Comportement des utilisateurs

Composant 36 : Le comportement de l'utilisateur d'un document spécifique peut entrer en ligne de compte.

Composant 37 : Il s'agit d'évaluer le nombre de clics vers un document dans des listes de résultats et le temps moyen passé par les utilisateurs sur ces documents.

Note : *L'évaluation du trafic et des taux de clics vers un document sont-ils des indicateurs de pertinence utilisés actuellement par Google ? Voir composants 15 et 16.*

Informations liées aux domaines

Composant 38 : Le calcul de la pertinence peut inclure les sites associés à un site donné et l'analyse de toutes les informations liées au domaine correspondant au domaine associé avec le document.

Composant 39 : Les documents sont en partie évalués en fonction de la "légitimité" des domaines associés aux documents.

Composant 40 : La date d'expiration du domaine, le "domain name server record" et le "name server associated with the domain" sont des critères permettant d'établir la légitimité d'un site associé.

Note : *On comprend mieux maintenant pourquoi Google a effectué une demande de "registrar" auprès de l'ICANN (voir <http://actu.abondance.com/2005-05/google->*

[registrar.php](#)). Le fait d'obtenir "à la source" ce type d'information représente un réel "plus" pour lui...

Données sur les classements préalables

Composant 41 : Le calcul de pertinence peut également prendre en compte les informations relatives à l'historique des classements préalables des documents (*prior ranking*).

Composant 42 : Google peut également calculer une quantité ou un taux de changements concernant le document pendant une période donnée (où le document se situait dans les classements préalables et comment il a évolué).

Composant 43 : Google, grâce à une prise en compte des classements préalables, change son calcul de la pertinence en fonction des résultats d'un site au fil du temps. Cela protège contre l'engouement momentané pour un site particulièrement efficace pour une recherche mais moins pour une autre.

Composant 44 : Les "pics" dans le classement d'un site dans les résultats de recherche sont également surveillés par Google pour savoir s'il s'agit d'une explosion "momentanée" ou d'un phénomène plus durable.

Note : L'analyse de l'historique du classement d'un site doit-elle permettre au moteur de définir son intérêt saisonnier (i.e. site d'achats de Noël) ou permanent ? L'identification des "pics" a-t-elle également pour fonction de repérer les sites "spammeurs" ?

Données utilisateurs

Composant 45 : Les données provenant des utilisateurs peuvent être enregistrées et analysées pour l'établissement du classement.

Composant 46 : Ces données incluent les listes de favoris, les fichiers temporaires et les fichiers cache.

Composant 47 : Les informations sur la vitesse à laquelle un site est ajouté ou retiré des données maintenues par les utilisateurs (comme les favoris) peuvent être prise en compte pour le calcul de la pertinence.

Note : Google a-t-il l'intention d'utiliser ses toolbars pour glaner des informations sur les utilisateurs ? Peut-on imaginer une interaction entre la recherche Web et l'outil "Google Desktop Search" ? D'autre part, la disponibilité d'outils comme Picasa, Hello, la "Google toolbar" ou le "Google Web Accelerator" n'ont-ils pas pour finalité de mieux comprendre comment les internautes utilisent le Web, afin d'appliquer ces comportements à son algorithme de pertinence ?

Profils de croissance des ancres des liens hypertextes

Composant 48 : Les calculs de pertinence peuvent s'appuyer sur la croissance du nombre d'ancres associées à un ou plusieurs liens hypertextes renvoyant vers un document.

Note : Il semblerait que ce contrôle du nombre d'ancres et de liens renvoyant vers un document explique lui aussi en partie l'effet "Sandbox", en pénalisant les sites qui ont une politique de référencement consistant à multiplier rapidement les liens vers leurs contenus.

Liens "extérieurs aux thèmes abordés"

Composant 49 : Pour déterminer la pertinence, les liens "éloignés" du sujet (*linkage of independent peers*) sont analysés pour définir le lien qui les relie au document pointé.

Utilisation des thèmes / sujets

Composant 50 : Les thèmes / sujets des documents (*Document topics*) peuvent être inclus dans le calcul de la pertinence, ce qui implique d'utiliser une technique d'extraction de thèmes.

Note : *Par quel biais le moteur peut-il définir la thématique d'une page ? Le rachat de sociétés comme Kaltix ou Applied Semantics, travaillant notamment dans la "compréhension" du Web, est-il lié à ce type de projets ?*

Identification des documents pertinents

Composant 51 : Le système de classement évaluera la pertinence des documents trouvés pour une requête donnée.

Pluralité des données historiques

Composant 52 : Pour évaluer la pertinence des documents, Google pourrait utiliser un ensemble de données historiques associées à ces documents. Tous les éléments de données historiques seront analysés par un système global.

Données historiques

Composant 53 : Le système de "scoring" comprend un composant historique configuré pour obtenir une ou plusieurs catégories de données historiques associées au document et un composant de "ranking" configuré pour calculer la pertinence du document.

Ranking des documents liés

Composant 54 : Google peut également prendre en compte, dans son évaluation, un critère d'altération ou d'ancienneté des liens pointant vers les documents. Cela signifie que des liens "récents" et des liens "périmés" seront pris en compte différemment.

Composant 55 : Les pages ne comportant aucun lien ne seront pas analysées sous cet angle.

Composant 56 : Les ancres pourront être incluses dans les données concernant les liens.

Composant 57 : Le classement sera donc basé sur les liens et les ancres.

Composant 58 : La longévité et la stabilité des données concernant les liens sont pris en compte pour déterminer le classement.

Composant 59 : Une pénalité sera infligée par Google si le critère de longévité montre que les liens ont une durée de vie courte. Au contraire, une récompense sera décernée pour les liens qui ont une longue durée de vie.

Composant 60 : Une pénalité sera également infligée lorsque les documents liants sont considérés comme "périmés" et un bonus sera accordé lorsque les pages pointant vers le site ont fréquemment actualisées.

Note : *Pour son calcul de pertinence, le moteur privilégie-t-il les documents disposant de backlinks sur des sites fréquemment actualisés ?*

Composant 61 : Le taux de renouvellement des liens (*link churn*) est évalué.

Composant 62 : Cette évaluation se fait sur la base des changements effectués sur les sites pointés, leurs ancres...

Composant 63 : "La multiplication abusive des liens", au delà d'un certain seuil, peut entraîner une pénalité.

Conclusion

Cette demande de brevet apporte des informations sur les critères qui sont jugés importants par Google pour déterminer la pertinence d'un document. Ces critères sont peut être déjà utilisés dans l'algorithme du moteur ou ils risquent de l'être dans les années à venir. Ils tendent pour certains à confirmer plusieurs points suspectés depuis longtemps quant au fonctionnement de Google (notamment sur l'effet "Sandbox" ou la prise en compte des taux de clics vers les documents pour déterminer la pertinence de ces documents...).

Cette demande de brevet témoigne en tout état de cause de la volonté de Google de protéger sa technologie pour éviter qu'elle ne soit utilisée prochainement par ses concurrents. Le moteur s'attend-il à un durcissement à venir de la concurrence sur la recherche d'information (ce en quoi il n'a certainement pas tort ;-)) ? A noter que Google vient de déposer une autre demande de brevet baptisée "*Systems and methods for improving the ranking of news articles*" au sujet de son moteur Google News. Cette demande, qui vise à breveter une sorte de "Newsrank" ou "Trustrank", fera l'objet d'une autre analyse dans cette lettre (voir <http://actu.abondance.com/2005-18/newsrank.php>).

Quelques liens...

Brevet

Information retrieval based on historical data - United States Patent Application (20050071741) - 31/03/2005

<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&co1=AND&d=PG01&s1=20050071741&OS=20050071741&RS=20050071741>qui>

Articles

Google Patent Analysis

<http://www.wolf-howl.com/2005/04/google-patent-analysis.html>

Google's Patent: Information Retrieval Based on Historical Data

<http://www.socengine.com/seo/guide/google-historical-data-patent.html>

Google United - Google Patent Examined

<http://www.webpronews.com/insiderreports/searchinsider/wpn-49-20050408GoogleUnitedGooglePatentExamined.html>