

Nous en avons parlé sur le site Abondance (<http://actu.abondance.com/2003-09/google-brevet.html>), Google a dernièrement obtenu un brevet auprès de l'ESPTO, organisme gérant les brevets aux Etats-Unis. L'occasion nous a semblé intéressante de faire le point, mois après mois, sur les différents brevets détenus par les acteurs importants de la recherche d'information sur le Web, car ceux-ci pourraient s'avérer très importants, stratégiquement parlant, à l'avenir. A tout seigneur tout honneur, nous commençons donc ce mois-ci avec Google. Selon nos recherches, la base de données de l'USPTO contient deux brevets spécifiquement au nom de Google :

"Ranking search results by reranking the results based on local inter-connectivity"

Brevet # 6,526,440

Appl. No.: 771677

30 janvier 2001

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=2&f=G&l=50&co1=AND&d=ptxt&s1=google.ASNM.&OS=AN/google&RS=AN/google>

Ce brevet présente un concept venant compléter celui du PageRank, utilisé actuellement par le moteur de recherche comme critère majeur de "ranking", par une pondération du classement initial des pages (qui est effectué indépendamment de la requête, en combinant des informations lexicales et le PageRank), par un deuxième classement local, dans le contexte de la requête. Il s'agit d'améliorer le score de pertinence d'un document en analysant l'inter-connectivité de ce dernier à l'intérieur d'un sous-index de documents reconnus eux-mêmes comme pertinents. Dans un premier temps, un sous-index "local" est créé, regroupant des documents répondant globalement à une requête par la présence du mot demandé dans leur texte. A l'intérieur de cet index local, un "score local" est calculé pour chaque page en tenant compte de l'interconnectivité des pages entre elles. Le score de pertinence, par rapport à la requête demandée initialement, est alors recalculé en fonction du "score local" obtenu par l'analyse de l'interconnexion des liens. Pour simplifier, la méthode proposée utilise 2 "PageRank" : l'un est calculé au niveau de l'index global, l'autre au niveau d'un index local, regroupant les pages répondant à la requête. Le classement final utilise une combinaison des 2 PageRank.

"Methods and apparatus for using a modified index to provide search results in response to an ambiguous search query"

Brevet # 6,529,903

Appl. No.: 748833

26 décembre 2000

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=/netahtml/search-adv.htm&r=1&p=1&f=G&l=50&d=ptxt&S1=google.ASNM.&OS=an/google&RS=AN/google>

Ce brevet a trait notamment aux interrogations d'un index à l'aide d'un terminal "non adapté" à une interrogation alphanumérique. Exemple : un clavier de téléphone portable. Actuellement, pour saisir une phrase (par exemple, pour taper un SMS), l'utilisateur doit éventuellement frapper successivement la même touche pour arriver à la bonne lettre. Exemple donné par le brevet : pour taper "ben smith", le possesseur d'un téléphone GSM doit taper :

b : 22

e : 33

n : 66

espace : 0

s : 7777

m : 6

i : 444

t : 8

h : 44

Le but du brevet est de décrire une solution qui permettrait de décrire, au prix de quelques approximations, chaque lettre uniquement par le chiffre qui la représente, et donc ici :

b : 2

e : 3

n : 6

espace : 0

s : 7
m : 6
i : 4
t : 8
h : 4

L'avantage est qu'on ne saisit que 9 chiffres (236076484) au lieu de 18 (223366077776444844) dans le cas précédent.

Le but est de faire en sorte que, statistiquement, le mot ainsi obtenu, même s'il est imprécis ("car" est défini, de façon numérique, de la même façon que "bar"), soit comparé à un index, transformé d'alphanumérique en numérique de la même façon, afin de pouvoir effectuer des recherches sur le Web rapidement. Si "bar" et "car" sont décrits de la même façon et donneront des résultats approximatifs, "wine", par exemple, ne propose aucune ambiguïté (9463) selon Google. Le moteur de recherche indique, dans le brevet, que cette méthode de recherche peut être étendu à d'autres formats, notamment phonétique et audiovisuel.

Notons que **Krishna Bharat** (photo ci-contre), - transfuge de Compaq -, inventeur notamment du premier brevet présenté dans cet article, en détient un certain nombre d'autres (<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HTOFF&p=1&u=%2Fmetahtml%2Fsearch-bool.html&r=0&f=S&l=50&TERM1=Bharat+krishna&FIELD1=INZZ&co1=AND&TE RM2=&FIELD2=&d=ptxt>) :



Numéro du brevet	Titre	Pour le compte de la société :
6,526,440	Ranking search results by reranking the results based on local inter-connectivity (voir ci-dessus)	Google
6,487,555	Method and apparatus for finding mirrored hosts by analyzing connectivity and IP addresses	AltaVista (voir article dans une prochaine lettre)
6,411,952	Method for learning character patterns to interactively control the scope of a web crawler	Compaq (voir article dans une prochaine lettre)
6,321,220	Method and apparatus for preventing topic drift in queries in hyperlinked environments	AltaVista (voir article dans une prochaine lettre)
6,286,006	Method and apparatus for finding mirrored hosts by analyzing urls	AltaVista (voir article dans une prochaine lettre)
6,195,698	Method for selectively restricting access to computer systems	Compaq (voir article dans une prochaine lettre)
6,141,760	System and method for generating unique passwords	Compaq (voir article dans une prochaine lettre)
6,112,203	Method for ranking documents in a hyperlinked environment using connectivity and selective content analysis	AltaVista (voir article dans une prochaine lettre)
5,815,709	System and method for generating identifiers for uniquely identifying object types for objects used in processing of object-oriented programs and the like	San Microsystems
5,621,546	Method and apparatus for generating identifiers	Xerox Corporation

	with output color control	
--	----------------------------------	--

Sanjay Ghemawat, co-inventeur d'un des brevets détenus par Google, détient, pour sa part, les brevets suivants :

Numéro du brevet	Titre	Pour le compte de la société :
6,529,903	Methods and apparatus for using a modified index to provide search results in response to an ambiguous search query (voir ci-dessus)	Google
6,253,370	Method and apparatus for annotating a computer program to facilitate subsequent processing of the program	Compaq (voir article dans une prochaine lettre)
6,073,135	Connectivity server for locating linkage information between Web pages	AltaVista (voir article dans une prochaine lettre)
5,956,758	Method for determining target address of computed jump instructions in executable programs	Digital Equipment (voir article dans une prochaine lettre)
5,796,939	High frequency sampling of processor performance counters	Digital Equipment (voir article dans une prochaine lettre)

On peut s'apercevoir, ici, que Google a débauché pas mal d'ingénieurs de chez AltaVista et ses propriétaires successifs (Digital, Compaq), notamment ceux qui détenaient des brevets intéressants ;-)

Notons également que **Sergey Brin**, cofondateur de Google, détient le brevet "**Method and apparatus for dynamically counting large itemsets**" (numéro 6,185,559) (<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=2&f=G&l=50&co1=AND&d=ptxt&s1='Brin+Sergey'.INZZ.&OS=IN/'Brin+Sergey'&RS=IN/'Brin+Sergey'>) pour le compte d'Hitachi Etats-Unis.

Larry Page, autre co-fondateur de Google, est propriétaire du brevet dénommé "**Method for node ranking in a linked database**" (numéro 6,285,999), qui décrit le principe du PageRank pour The Board of Trustees of the Leland Stanford Junior University (<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=1&f=G&l=50&co1=AND&d=ptxt&s1='Page+Lawrence'.INZZ.&OS=IN/'Page+Lawrence'&RS=IN/'Page+Lawrence'>).

Ce brevet sera plus explicitement exploré dans un article de la lettre "Recherche & Référencement" du mois prochain.

Google a également 2 autres demandes de brevets en attente :

"Methods and apparatus for employing usage statistics in document retrieval"

Demande # 20020123988

5 septembre 2002

URL : <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=2&f=G&l=50&co1=AND&d=PGO1&s1=google.AS.&OS=AN/google&RS=AN/google>

Cette demande de brevet décrit un classement des pages basé sur des critères "classiques" pour un moteur de recherche, auxquels ont été ajoutés un critère basé sur des statistiques ("usage statistics"), sur la façon dont le document a été utilisé : par exemple le nombre de personnes qui l'ont déjà consulté, le nombre d'affichages de la page, quel que soit le nombre de visiteurs, et ce éventuellement sur une durée prédéfinie. Pour les amateurs de mathématiques, le score de fréquence de visite d'une page est donné par la formule $\log_2(1 + \log(VF)) / \log(\text{MAXVF})$, où VF représente le nombre de fois où un document est affiché en un mois et MAXVF une constante fixée à 2000.

Ghemawat Sanjay, l'un des inventeurs du brevet, a également déposé une demande pour un brevet proche, nommé "**System and method for reorganizing data storage in accordance with usage frequency**" (demande 20020087563) pour la société californienne Pennie & Edmonds, LLP.

"Methods and apparatus for using a modified index to provide search results in response to an ambiguous search query"

Demande # 20020042791

11 avril 2002

URL : <http://appft1.uspto.gov/netacgi/nph->

[Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-](http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=3&f=G&l=50&co1=AND&d=PG01&s1=google.AS.&OS=AN/google&RS=AN/google)

[bool.html&r=3&f=G&l=50&co1=AND&d=PG01&s1=google.AS.&OS=AN/google&RS=AN/google](http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=3&f=G&l=50&co1=AND&d=PG01&s1=google.AS.&OS=AN/google&RS=AN/google)

Cette demande de brevet est connexe au brevet, déjà obtenu par Google, nommé "Methods and apparatus for using a modified index to provide search results in response to an ambiguous search query" (voir ci-dessus). Il décrit la façon dont un index alphanumérique "conventionnel" peut être converti en un index numérique afin d'être questionné par un terminal disposant d'un clavier simplifié, basé sur des touches numériques, comme un téléphone par exemple.