

Un point sur les brevets : AltaVista

Nous continuons notre série d'articles sur les brevets en nous penchant ce mois-ci sur ceux que détient le moteur de recherche AltaVista.

Sur le site de l'USPTO (<http://www.uspto.gov/patft/index.html>), lorsque nous saisissons le terme "altavista" dans le champ "Assignee name" ("propriétaire du brevet"), nous obtenons 13 enregistrements. Les voici :

Numéro du brevet	Titre
6,349,296	Method for clustering closely resembling data objects
6,332,144	Technique for annotating media
6,321,265	System and method for enforcing politeness while scheduling downloads in a web crawler
6,321,220	Method and apparatus for preventing topic drift in queries in hyperlinked environments
6,317,741	Technique for ranking records of a database
6,311,189	Technique for matching a query to a portion of media
6,275,827	Technique for processing data
6,266,657	Technique for indexing data in a network
6,230,158	Method for indexing duplicate records of information of a database
6,230,155	Method for determining the resemining the resemblance of documents
6,219,671	Technique for storing data information within a network
6,138,113	Method for identifying near duplicate pages in a hyperlinked database
6,112,203	Method for ranking documents in a hyperlinked environment using connectivity and selective content analysis

Lorsque l'on saisit "alta vista" (en deux mots) dans le formulaire, on obtient encore 10 nouveaux brevets :

Numéro du brevet	Titre
6,560,600	Method and apparatus for ranking Web page search results
6,513,032	Search and navigation system and method using category intersection pre-computation
6,487,555	Method and apparatus for finding mirrored hosts by analyzing connectivity and IP addresses
6,377,984	Web crawler system using parallel queues for queing data sets having common address and concurrently downloading data associated with data set in each queue
6,351,755	System and method for associating an extensible set of data with documents downloaded by a web crawler
6,301,614	System and method for efficient representation of data set addresses in a web crawler
6,286,006	Method and apparatus for finding mirrored hosts by analyzing urls
6,269,362	System and method for monitoring web pages by comparing generated abstracts
6,263,364	Web crawler system using plurality of parallel priority level queues having distinct associated download priority levels for prioritizing document downloading and maintaining document freshness
6,073,135	Connectivity server for locating linkage information between Web pages

On peut d'ailleurs s'étonner que le nom de l'inventeur d'un brevet ne soit pas normalisé et que le moteur de recherche AltaVista soit ainsi présent dans la base de données de l'USPTO sous deux orthographes différentes.

D'autre part, lors du rachat d'AltaVista, Overture a communiqué sur "les 58 brevets détenus" par ce moteur de recherche. Or, il semble donc en manquer 25 à l'appel. Sont-ils déposés sous un autre nom ? (il en existe de nombreux sous les noms de Digital et de Compaq, propriétaires successifs du moteur, mais dans ce cas, les brevets appartiennent-ils à AltaVista ?) Sont-ils déposés dans

d'autres pays ? Voici la réponse de Susan Marsch, Directeur juridique worldwide d'AltaVista :
"When patents are assigned, the assignments don't show up online at the PTO office. You'd actually have to go to the PTO or look at the Assignment agreement from the CMGI/Compaq acquisition. So a lot of the DEC issued patents were assigned and probably still show up in DEC or Compaqs name, but we have all the appropriate paperwork to establish ownership, as Overture can attest to."

Comment savoir, alors, quels sont les 25 brevets appartenant à AltaVista et déposés sous le nom de Digital ou de Compaq ? Nous recherchons, actuellement, des informations à ce sujet et vous tiendrons au courant dans les mois qui viennent.

Les 23 brevets "officiels" d'AltaVista peuvent cependant être visualisés à l'adresse (en utilisant la syntaxe OU, proposée par le site, entre les deux orthographes) :
<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=0&f=S&l=50&TERM1=altavista&FIELD1=ASNM&co1=OR&TERM2=alta+vista&FIELD2=ASNM&d=ptxt>

Examinons rapidement ces brevets, classés par ordre chronologique (du plus ancien au plus récent) :

* Brevet numéro 6 269 362 : **System and method for monitoring web pages by comparing generated abstracts.**

URL : [http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=8&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista"&S=AN/'alta+vista"](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=8&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista)

Date : 19 décembre 1997.

Description : Ce brevet explique comment il est possible de comparer le contenu de plusieurs pages en en extrayant des résumés ou des extraits et en comparant leurs contenus à des intervalles de temps différents.

* Brevet numéro 6 073 135 : **Connectivity server for locating linkage information between Web pages.**

URL : [http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=10&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista"&RS=AN/'alta+vista"](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=10&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista)

Date : 10 mars 1998.

Description : Ce brevet décrit la façon dont AltaVista prend en compte l'indice de popularité des documents web au travers de leur interconnectivité, sur la base d'un "serveur de connectivité" fournissant des informations sur la façon dont sont liées entre elles les pages du Web. Le brevet décrit la façon dont ce "serveur de connectivité" a été imaginé par les ingénieurs du moteur de recherche.

* Brevet numéro 6 112 203 : **Method for ranking documents in a hyperlinked environment using connectivity and selective content analysis.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=13&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 9 avril 1998.

Description : Ce brevet décrit le système d'indice de popularité utilisé par AltaVista. Il complète l'algorithme de Kleinberg qui définit les pages dites "Authority" (nombreux liens entrants) et les "Hubs" (de nombreux liens sortant vers les pages de type "authority").

* Brevet numéro 6 138 113 : **Method for identifying near duplicate pages in a hyperlinked database.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=12&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 10 août 1998.

Description : Ce brevet décrit une méthode pour dédoublonner un index en utilisant l'interconnectivité des pages le composant, et notamment le nombre de liens sortant de chaque page comparée.

* Brevet numéro 6 230 155 : **Method for determining the resemining the resemblance of documents.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=10&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 23 novembre 1998.

Description : Un autre brevet (voir les brevets 6 230 158, 6 138 113 et 6 349 296) décrivant une méthodologie de dédoublonnage d'index. La méthode est basée sur la comparaison de résumés ("sketches") des documents plus que sur une comparaison brute du contenu complet des pages.

* Brevet numéro 6 332 144 : **Technique for annotating media.**

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=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 3 décembre 1998.

Description : Ce brevet propose une technique d'annotation de fichiers multimédia (vidéo, audio) afin de rechercher plus facilement ce type de fichier sur la base de son contenu.

* Brevet numéro 6 311 189 : **Technique for matching a query to a portion of media.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=6&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 3 décembre 1998.

Description : Ce brevet complète le brevet numéro 6 332 144 et propose une méthodologie pour faire correspondre la requête d'un utilisateur à une portion de fichier multimédia (audio/vidéo notamment).

* Brevet numéro 6 219 671 : **Technique for storing data information within a network.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=11&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 3 décembre 1998.

Description : Ce brevet décrit une méthode de stockage d'information sur un réseau de serveurs. Le brevet se rapporte plutôt au monde multimédia (fichiers audio / vidéo / image). Le réseau distribué peut être représenté par un intranet ou une partie de l'Internet.

* Brevet numéro 6 275 827 : **Technique for processing data.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=6&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 3 décembre 1998.

Description : Ce brevet se rapporte également au traitement de documents multimedia (audio, vidéo, image). Il décrit une méthode permettant de "découper" un document (par exemple une vidéo) en différentes séquences qui pourront par la suite faire l'objet de recherches spécifiques.

* Brevet numéro 6 321 220 : **Method and apparatus for preventing topic drift in queries in hyperlinked environments.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=4&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 7 décembre 1998.

Description : Ce brevet décrit une méthode pour classer des documents en fonction d'une thématique précise tout en limitant les dérives possibles liées à l'utilisation d'une méthode basée sur l'interconnectivité des documents. Une requête, associée à une thématique, est fournie au moteur. Un graphe d'interconnectivité est alors bâti entre les pages répondant à la requête. Le graphe est utilisé pour vérifier si les résultats sont bien en correspondance avec la thématique demandée. Les documents ne répondant pas à la thématique sont éliminés. Les autres sont à nouveau triés par ordre de pertinence.

- * Brevet numéro 6 487 555 : **Method and apparatus for finding mirrored hosts by analyzing connectivity and IP addresses.**

URL : [http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=3&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/"alta+vista"&RS=AN/"alta+vista"](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=3&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/)

Date : 7 mai 1999.

Description : Ce brevet décrit une méthode pour détecter les sites miroirs (sites identiques ou quasi-identiques dans leurs contenus mais ayant des adresses IP et noms de domaines différents). Les méthodes décrites se basent notamment sur l'interconnectivités des pages du Web.

- * Brevet numéro 6 286 006 : **Method and apparatus for finding mirrored hosts by analyzing urls.**

URL : [http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=7&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/"alta+vista"&RS=AN/"alta+vista"](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=7&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/)

Date : 7 mai 1999.

Description : Ce brevet décrit une méthodologie de reconnaissance de sites miroirs en utilisant des informations incluses dans un large échantillon de pages web, et notamment l'interconnectivité entre les documents.

- * Brevet numéro 6 230 158 : **Method for indexing duplicate records of information of a database.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=9&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 19 octobre 1999.

Description : Ce brevet décrit une méthode pour dédoublonner des documents identiques stockés dans un index, chacun des documents recevant une "empreinte digitale" destinée à le caractériser et à identifier d'autres documents proches qui seront ainsi éliminés.

- * Brevet numéro 6 266 657 : **Technique for indexing data in a network.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=8&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 26 octobre 1999.

Description : Ce brevet décrit une méthode pour indexer des données dans un réseau distribué contenant plusieurs machines, plusieurs serveurs. Le brevet se rapporte plutôt au monde multimédia (fichiers audio / vidéo / image). Le réseau distribué peut être représenté par un intranet ou une partie de l'Internet.

- * Brevet numéro 6 513 032 : **Search and navigation system and method using category intersection pre-computation.**

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='alta+vista'.ASNM.&OS=AN/"alta+vista"&RS=AN/"alta+vista"](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='alta+vista'.ASNM.&OS=AN/)

Date : 28 octobre 1999.

Description : Ce brevet décrit une méthode pour classifier de façon automatique des documents en fonction des thématiques ("report keys") dont il est question dans son contenu.

- * Brevet numéro 6 321 265 : **System and method for enforcing politeness while scheduling downloads in a web crawler.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahhtml/search-bool.html&r=3&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 2 novembre 1999.

Description : Ce brevet propose une méthode de gestion des files d'attente dans le cadre du crawl du Web afin de constituer l'index d'un moteur.

- * Brevet numéro 6 377 984 : **Web crawler system using parallel queues for queuing data sets having common address and concurrently downloading data associated with data set in each queue.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahhtml/search-bool.html&r=4&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista'&RS=AN/'alta+vista'>

Date : 2 novembre 1999.

Description : Ce brevet décrit une méthode pour mieux gérer les files d'attente utilisées par un spider pour télécharger les pages du Web afin de générer l'index d'un moteur de recherche.

- * Brevet numéro 6 263 364 : **Web crawler system using plurality of parallel priority level queues having distinct associated download priority levels for prioritizing document downloading and maintaining document freshness.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahhtml/search-bool.html&r=9&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista'&RS=AN/'alta+vista'>

Date : 2 novembre 1999.

Description : Ce brevet complète le brevet 6 351 755 et donne des indications sur la façon dont les différentes files d'attente sont gérées pour mener au mieux une phase de crawl de la part des spiders d'un moteur de recherche, et notamment les niveaux de priorité de chaque file d'attente, en fonction des délais de mise à jour des documents (une page mise à jour quotidiennement doit pouvoir être crawlée plus souvent qu'un document mis à jour en moyenne de façon mensuelle).

- * Brevet numéro 6 351 755 : **System and method for associating an extensible set of data with documents downloaded by a web crawler.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahhtml/search-bool.html&r=5&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista'&RS=AN/'alta+vista'>

Date : 2 novembre 1999.

Description : Ce brevet décrit une méthodologie utilisée par les spiders, utilisant une structure appelée "The Frontier" pour mettre en file d'attente les urls des pages à crawler. Cet outil stocke également des informations complémentaires sur le types de données à stocker (taille, date et heure de crawl, date et heure de dernière modification, etc.), leur date de validité, les niveaux de priorité des différentes files d'attente utilisées, etc.

- * Brevet numéro 6 301 614 : **System and method for efficient representation of data set addresses in a web crawler.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahhtml/search-bool.html&r=6&f=G&l=50&co1=AND&d=ptxt&s1='alta+vista'.ASNM.&OS=AN/'alta+vista'&RS=AN/'alta+vista'>

Date : 2 novembre 1999.

Description : Ce brevet décrit une méthodologie de gestion des caches et des adresses de documents et de liens lors du téléchargement de pages web par un spider.

- * Brevet numéro 6 317 741 : **Technique for ranking records of a database.**

URL : <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=5&f=G&l=50&co1=AND&d=ptxt&s1=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 7 août 2000.

Description : Ce brevet décrit la méthodologie du moteur de recherche pour établir un "ranking" des documents à classer lors de la saisie d'une requête par un internaute, en se basant notamment sur le nombre d'occurrences des mots recherchés à l'intérieur des documents de l'index.

* Brevet numéro 6 349 296 : **Method for clustering closely resembling data objects.**

URL : <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=altavista.ASNM.&OS=AN/altavista&RS=AN/altavista>

Date : 21 août 2000.

Description : Ce brevet décrit une méthode pour dédoublonner l'index d'un moteur de recherche, notamment pour en éliminer des pages très ressemblantes mais implémentées sur des serveurs différents (techniques de pages alias et de mirroring). Le but étant d'obtenir une copie unique d'une page web donnée et d'en éliminer toutes les autres versions.

* Brevet numéro 6 560 600 : **Method and apparatus for ranking Web page search results.**

URL : <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='alta+vista'.ASNM.&OS=AN/'alta+vista'&RS=AN/'alta+vista'>

Date : 25 octobre 2000.

Description : Ce brevet décrit un algorithme de classement basé sur l'utilisation de plusieurs matrices, donnant une image de la pertinence de pages web en relation à une requête. Des matrices donnant, par exemple, une idée de la qualité de certaines pages et du fait que d'autres puissent être indésirables peuvent également être utilisées.

Deux demandes semblent également être en cours de traitement auprès de l'USPTO de la part d'AltaVista :

* Dossier numéro 20020049753 : **Technique for deleting duplicate records referenced in an index of a database.**

URL : <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=altavista.AS.&OS=AN/altavista&RS=AN/altavista>

Date : 25 avril 2002.

Description : Le nom du dossier est explicite : il s'agit d'une technique qui permet de dédoublonner l'index d'un moteur de recherche (encore une !).

* Dossier numéro 20010051958 : **Technique for locating an item of interest within a stored representation of data.**

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=PG01&s1=altavista.AS.&OS=AN/altavista&RS=AN/altavista>

Date : 13 décembre 2001.

Description : La demande de brevet touche au domaine du multimédia (vidéo, audio, image). Il fait référence à une technique permettant de trouver des documents multimédia sur la base d'annotations appliquées à chaque document.

Conclusion

Les brevets déposés (ou en attente) par AltaVista ont donc trait aux domaines suivants :

La **comparaison** de documents : 1 brevet.

L'**analyse** et le **ranking** de résultats : 3 brevets.

La gestion des **files d'attente** des *spiders* : 5 brevets.

L'indice de popularité, l'analyse de l'**interconnectivité** des documents : 3 brevets.

La **déduplication** / dédoublement d'un index et détection de sites miroirs : 6 brevets et 1 demande en cours.

La recherche sur des documents **multimédias** (audio, vidéo, image) : 5 brevets et 1 demande en cours.

Notons également que nous n'avons trouvé aucun brevet déposé au nom de CMGI, actuel propriétaire d'AltaVista et que le dernier brevet accepté par l'USPTO au nom du moteur date du mois d'octobre 2000, soit il y a plus de deux ans et demi...

Pour l'anecdote (quoi que...), terminons cet article sur le fait que Michael Burrows, l'un des concepteurs du moteur AltaVista, travaille aujourd'hui... Chez Microsoft, société qui semble grandement oeuvrer à la mise en place d'une technologie "moteur". Or, le nom de Michael Burrows est associé à... 46 brevets (<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=0&f=S&l=50&TERM1=Burrows:+Michael&FIELD1=INZZ&co1=AND&TERM2=&FIELD2=&d=ptxt>), la plupart déposés pour le compte de Digital, ancien propriétaire d'AltaVista. On peut quand même noter quelques homonymes dans cette liste (assez peu cependant).