

Google dépose un brevet sur un 'moteur de similarité'

Le **brevet 7.158.961** que le bureau américain des brevets (*U.S. Patent and Trademark Office*) vient d'accorder à Google porte sur le déploiement d'un '**moteur de similarité**' (*similarity-engine*).

L'une des difficultés des moteurs de recherche provient de la multiplication des résultats extraits de leurs bases qui proposent les mêmes informations, parfois simplement présentées différemment mais au contenu identique ou très proche.

Difficile dans ces conditions de cataloguer l'information, et l'internaute ou l'utilisateur local se voient contraints de passer par ces documents souvent identiques ou similaires, ce qui entraîne souvent un haut niveau de frustration..

Google a donc développé un moteur dont les algorithmes créent et calculent des sommes et des différences exprimées en vecteurs. A l'aide de ces 'ensembles traités', le brevet évoque des '*ashes*' et des '*sketches*', le moteur est capable de comparer les différences dans le texte comme dans les images.

Le moteur de similarité part donc d'un objet, le vectorise, et compare le vecteur ainsi créé avec le vecteur d'un autre objet? Cette méthode selon Google pourra être utilisée en dehors des pages web, pour des documents textuels, des tableaux, des présentations, théoriquement tout type de représentation des données de l'entreprise.

Il sera difficile en revanche pour Google de faire prévaloir son brevet sur les projets et applications de 'moteur de similarité'. En effet, en dix ans l'office américain des brevets a validé une quinzaine de brevets sur le même thème, dont certains émanent d'IBM, Hitachi ou Visage.

A ce propos, le '*similarity-engine*' de Google a-t-il détecté ces similarités dans les bases de brevets ?