

Evri, un extraordinaire outil de recherche sémantique dans l'actualité

[Retour au sommaire de la lettre](#)

Evri est un outil de recherche dans l'actualité assez peu connu mais qui gagnerait à l'être par ses fonctionnalités avancées et surtout son langage de recherche évoluée... Au travers des quelques exemples décrits dans cet article et basés sur sa syntaxe de recherche ainsi que sur quelques astuces utilisant la taxonomie et les entités nommées, vous vous apercevrez vite de la puissance de cet outil en termes d'intelligence économique et de veille. Voici vraiment un moteur qui montre enfin ce que pourrait être cette révolution du web sémantique aussi promise qu'attendue...

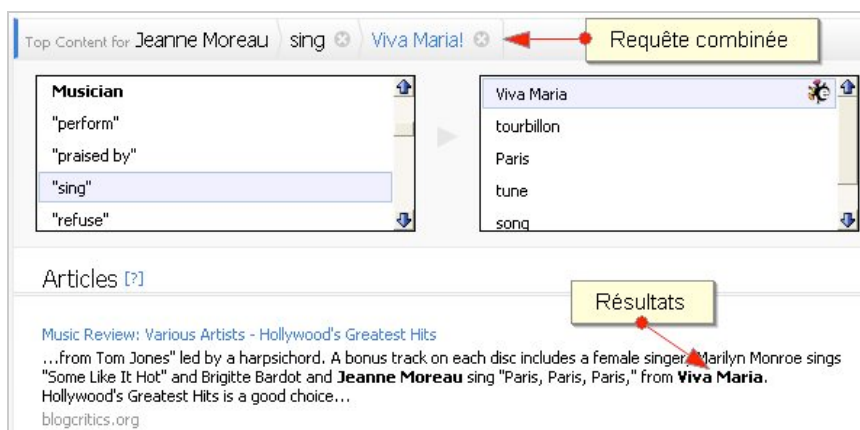
On parle beaucoup de la recherche sémantique et de ce qu'elle est sensée apporter à tous ceux qui cherchent de l'information au quotidien (c'est-à-dire presque tout le monde) : précision, qualité, puissance des requêtes, etc. On en parle beaucoup mais... on ne voit pas grand-chose venir.

Et pourtant, il existe un moteur de recherche qui nous donne une bonne idée des bénéfices que l'on peut en attendre. Ce moteur, quasiment inconnu et travaillant sur l'actualité, s'appelle Evri (<http://www.evri.com/>) et ses fonctionnalités "de base" suffisent à le rendre intéressant. Voyez vous-même :

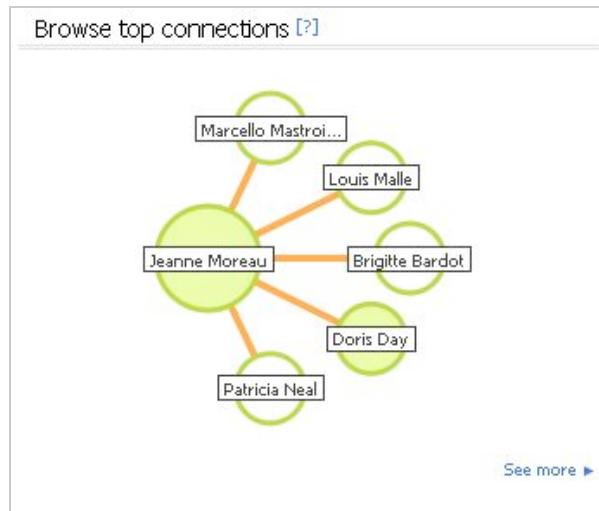
- Recommandations "as you type".
- Éléments tirés de la Wikipedia :



- Filtrage du contenu indexé par combinaison d'entités nommées :

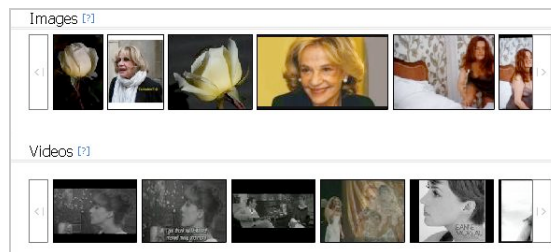


- Cartographie relationnelle dynamique:



- Courbes de tendances du mot-clé dans l'actualité.

- Images et vidéos :



Des fonctionnalités intéressantes, donc, mais à première vue, rien qui vienne vraiment démarquer Evri d'un Silobreaker, d'un Textmap ou d'un Newstin. Il faut en effet chercher un peu sur le blog de la société (<http://blog.evri.com/>) pour découvrir les fonctionnalités de recherche avancées (<http://www.evri.com/mainline-ui/jsp/index.jsf>). Et quelles fonctionnalités !

Grâce à un système à base de structuration sémantique de contenu, de détection d'entités nommées et de taxonomie, Evri a tout simplement mis au point un langage d'interrogation permettant à celui qui le maîtrise de fouiller l'actualité du web (et les archives indexées par Evri) de manière bien plus fine que tout ce que l'on connaissait jusqu'à maintenant. En tout cas bien plus que les autres moteurs de recherche d'actualité existants, étant entendu que les bases de données payantes (Lexis-Nexis, Factiva,...) proposent déjà des fonctionnalités avancées à base d'opérateurs toujours introuvables ailleurs (il n'y a guère qu'Exalead pour proposer la troncature). Toutefois même ces services sont selon nous dépassés par ce que propose Evri, comme nous allons le voir par la suite.

Des requêtes à base d'entités nommées

Pour comprendre ce que l'on peut faire d'Evri, il faut parler un tout petit peu de syntaxe d'interrogation, rien de compliqué en fait. Tout d'abord Evri distingue trois types d'items :

1. Source : le sujet qui nous intéresse (une personne, un pays, une entreprise...);
2. Action : n'importe quel verbe (source + action);
3. Cible : ce vers quoi l'action est dirigée.

En les combinant grâce à une syntaxe et des opérateurs spécifiques, vous allez créer entre eux une Relation. Attention, tout cela ne fonctionne pour l'instant que pour l'anglais.

Exemple 1 : Vous souhaitez obtenir des articles dans lesquels l'Inde accuse (*to accuse*) le Pakistan. Vous allez donc taper :

India>accuse>Pakistan

Les signes ">" sont les éléments de base de la syntaxe d'Evri.
 Vous obtenez alors 376 résultats classés par date :

Date	Count	Term 1	Term 2	Term 3
12/10/2008	[75]	India	accuse	Pakistan
11/30/2008	[9]	India / in	accuse	Pakistan
12/02/2008	[4]	India	accuse	Pakistani
12/06/2008	[1]	India	accuse	element
Hoax put Pakistanis on alert ...India has blamed Pakistan-based militants in the attacks but not the Pakistani state. Pakistan has denied official involvement, and there is widespread public anger that India accused Pakistani elements even as the attacks were unfolding. Related topic galleries: National Government, Asif Ali Zardari, Government, Defense, Diplomacy, Mumbai Terror Attacks, Fakes, Hoaxes, and Transcripts.				
12/04/2008	[6]	India / military	accuse	force
		India	accuse	Lashkar e-Taiba / group
11/27/2008	[1]	India	lodge	Pakistan
11/27/2008	[1]	complicity / India	accuse of	Pakistan / ISLAMABAD,
11/27/2008	[1]	complicity / India	accuse of	Pakistan
12/10/2008	[1]	India	accuse of	Pakistani / group / Lashkar-e-Taiba

Vous remarquez au passage qu'Evri a automatiquement utilisé des synonymes du verbe "accuse" ("accuse of", "lodge"). Il utilise de la même manière les différentes formes verbales d'un même mot, et si vous souhaitez conserver un verbe dans une forme précise il suffit de le mettre entre guillemets.

Notez enfin la petite croix qui vous donne accès à la prévisualisation des articles et, si nécessaire, à la page d'origine.

Mais Evri permet en fait beaucoup plus que cela. Mettons que vous souhaitez maintenant préciser un contexte pour les articles qui vous intéressent. Vous voulez qu'ils évoquent le groupe terroriste **Lashkar-e-Taiba**. Vous allez donc ajouter l'opérateur spécifique "CONTEXT CONTAINS":

India>accuse>Pakistan CONTEXT CONTAINS Lashkar-e-Taiba

et obtenir ainsi un nouvel ensemble de 15 résultats :

Date	Count	Term 1	Term 2	Term 3
--	[7]	India	accuse	Pakistan
12/01/2008	[3]	India	accuse	Lashkar-e-Taiba / Pakistan
12/02/2008	[2]	India	accuse	Lashkar-e-Taiba
12/03/2008	[1]	India	accuse	LeT / Lashkar e-Taiba / group
12/04/2008	[1]	India	lodge	protest / Pakistan
12/10/2008	[1]	India	accuse of	Pakistani / group / Lashkar-e-Taiba

Et ce n'est toujours pas tout puisque vous pouvez combiner ces éléments avec des opérateurs booléens classiques.

Si par exemple vous souhaitez utiliser le mot-clé "terrorist" pour élargir les résultats. Il vous suffit de taper :

India>accuse>Pakistan CONTEXT CONTAINS Lashkar-e-Taiba OR terrorists

Vous obtenez alors 77 résultats au lieu de 15.

India>accuse>Pakistan CONTEXT CONTAINS Lash					
Go help					
Results 1-77 of 77 for India>accuse>Pakistan CONTEXT CONTAINS Lashkar-e-Taiba OR terrorists					
expand all					
--	[+]	[65]	India	→	accuse → Pakistan
12/01/2008	[+]	[3]	India	→	accuse → Lashkar-e-Taiba / Pakistan
12/02/2008	[+]	[2]	India	→	accuse → Lashkar-e-Taiba
--	[+]	[2]	India	→	accuse → terrorist / Pakistan
--	[+]	[1]	India	→	accuse → Islamic / group
12/03/2008	[+]	[1]	India	→	accuse → Lashkar e-Taiba / group
12/04/2008	[+]	[1]	India	→	lodge → protest / Pakistan
12/10/2008	[+]	[1]	India	→	accuse of → group / Lashkar-e-Taiba
--	[+]	[1]	India	→	accuse of → The Government / Pakistan

Des requêtes utilisant la taxonomie

La détection d'entités nommées et leur intégration dans une taxonomie permet toutefois d'aller encore plus loin. Ce que nous avons vu jusque là nous a permis d'affiner une requête pour arriver à des résultats qui soient les plus précis possible. Nous avons maintenant la possibilité de faire l'inverse, c'est-à-dire de tester des requêtes pour vérifier si une information existe en réponse. Il va pour cela falloir nous familiariser maintenant avec la taxonomie d'Evri. Rien de compliqué, rassurez-vous.

Lorsque Evri découvre une nouvelle entité (personne, lieu...), il va la placer dans sa taxonomie (Paris → Ville, Vigo Mortensen → Acteur, ...). Vous allez donc pouvoir utiliser le nom d'une rubrique ("Ville", "Acteur") pour lancer une requête dans toutes les entités qu'elle contient (la taxonomie complète est disponible ici : <http://www.evri.com/mainline-ui/jsp/index.jsf>).

Exemple 2 (un peu plus léger) : vous souhaitez savoir quel acteur est en contrat avec quelle compagnie. Vous explorez la taxonomie et repérez les termes "Actor" et "Film_production_company" que vous allez organiser ainsi :

[Actor]>contract>[Film_production_company]

Vous lancez ainsi une requête pour tous les acteurs et toutes les sociétés de production recensés par Evri. Résultat :

[Actor]>contract>[Film_production_company]					
Go help					
Results 1-100 of 142 for [Actor]>contract>[Film_production_company]					
expand all					
09/26/2008	[+]	[1]	Patricia Kirkwood / 7-year	→	contract → offer / Metro-Goldwyn-Mayer
06/17/2008	[+]	[1]	Stan Winston	→	undertake → apprenticeship / Walt Disney Studios
10/29/2008	[+]	[14]	Robert Downey Jr.	→	sign → Marvel Studios
--	[+]	[8]	Catherine O'Hara / Allison Janney / Wanda Sykes	→	sign → DreamWorks
03/18/2008	[+]	[4]	Kirstie Alley	→	sign → Harpo Productions
--	[+]	[3]	Anika Noni Rose	→	sign → Disney
--	[+]	[3]	Bing Crosby	→	sign → Paramount Pictures
--	[+]	[3]	O'Hara	→	sign → RKO Studios
10/08/2008	[+]	[2]	Claudette Colbert	→	sign → Paramount Pictures
05/09/2008	[+]	[2]	David Chase	→	sign → Paramount Pictures
--	[+]	[2]	Frances Farmer	→	sign → Paramount Pictures
--	[+]	[2]	Jack Warner	→	sign → MGM
05/19/2008	[+]	[2]	John Cusack	→	sign → Columbia Pictures
--	[+]	[2]	Kevin Bacon	→	sign → Columbia Pictures

Simple et efficace !

Si vous souhaitez maintenant avoir les articles citant des montants de contrats, vous allez ajouter **PREP CONTAINS** à votre requête ainsi que l'élément de taxonomie [money]. En utilisant cette expression, vous supposez que la préposition dans laquelle se trouvent les éléments de votre

requête inclura aussi un montant chiffré. Comme par exemple dans l'hypothétique phrase suivante : "Jean-Claude Van Damme (Source) a signé un contrat (Action) avec Columbia (Cible) de 500 millions de \$ pour 5 ans".

Exemple 3 :

[Actor]>contract>[Film_production_company] PREP CONTAINS [money]

Résultat :

Dernier exemple, vous souhaitez connaître les sociétés qui licencient en Chine. Vous allez donc taper :

[Company]>dismiss OR fire>worker OR employee PREP CONTAINS China

Résultat :

Nous aimerions vous dire que c'est tout (car cela fait tout de même pas mal de choses à assimiler) mais ce n'est pourtant pas le cas. Il existe encore quelques opérateurs et syntaxes à découvrir mais nous n'en citerons qu'un. Il s'agit en fait plutôt d'une astuce que sa simplicité rend géniale. Nous avons vu en effet que le signe ">" était l'élément de syntaxe de base d'Evri. Or ce signe est aussi un symbole opératoire qui indique le sens de la relation, sens que l'on peut inverser. Pour être plus clair, imaginons que vous vouliez savoir quelles sont les personnes qui déclarent ne pas aimer Obama. Vous allez taper :

[people]>hate>obama

Si maintenant vous inversez les symboles :

[people]<hate<obama

Vous obtenez tout simplement la liste des personnes qu'Obama n'apprécie pas. Et si vous mêlez les symboles de cette manière :

[people]<>hate<>obama

Vous obtenez les résultats cumulés des deux précédentes requêtes.

Bien sur le rêve serait maintenant qu'Evri propose des flux Rss pour chacune des requêtes complexes qu'il permet d'effectuer. Mais n'oublions pas qu'il est encore en phase bêta, et d'ailleurs avec un tel degré d'innovation, on ne peut qu'être optimiste.

Quoi qu'il en soit, Evri est réellement un moteur qui montre enfin ce que pourrait être cette révolution du web sémantique aussi promise qu'attendue. Un outil qu'il faudra sans aucun doute suivre avec attention dans les mois et les années qui viennent...

Christophe Deschamps

Consultant et formateur en gestion de l'information.

Responsable du blog Outils Froids (<http://www.outilsfroids.net/>)

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://abonnes.abondance.com/blogpro/2008/12/evri-un-extraordinaire-outil-de.html>