

Silobreaker, un étonnant outil de veille

[Retour au sommaire de la lettre](#)

Silobreaker est un service gratuit d'agrégation de dépêches d'actualité bien plus puissant que des outils comme Google News ou Yahoo! News. Il pourrait vite devenir un outil majeur de veille, tant est grande sa qualité de services et la quantité de sources indexées. Voici un tour d'horizon de ses principales fonctionnalités...

Silobreaker (<http://beta.silobreaker.com/>) est un service d'agrégation d'actualités en phase bêta comparable à Google News ou Yahoo! News. Ou plutôt non, il est incomparablement plus puissant et plus intéressant en terme de fonctionnalités que ces deux-là réunis, ne serait-ce que par la quantité de sources ouvertes traitées puisqu'il en intègre plus de 9 000 en anglais, soit le double de Google News US.

Silobreaker se présente comme un portail classique et requiert une inscription (gratuite) pour profiter de certaines options de personnalisation.

Vous pouvez l'utiliser comme n'importe quel portail généraliste, c'est-à-dire en choisissant d'explorer les nombreuses catégories qu'il propose, au total environ 70 classées dans 3 catégories principales : « Global Issue », « Science & Technology », « Business », ainsi qu'une quatrième proposant un classement de l'information par pays.

Vous pouvez également choisir de lancer une requête dans le moteur de recherche de Silobreaker qui vous amènera sur une page de résultats parfaitement structurée. En fait Silobreaker traite l'actualité en temps réel et les pages qu'il propose, que ce soit en naviguant ou en lançant une recherche, sont des pages générées dynamiquement. Pour arriver à ce résultat il utilise une plateforme technique capable de traiter des contenus structurés, non-structurés, numériques, audio/video, mais aussi des métadonnées, des thésaurus et des taxonomies. Cette masse de données passe ensuite par une quadruple « moulinette » :

- Indexation.
- Extraction d'entités : reconnaissance de noms de personne, d'organisations, d'entreprises.
- Analyse des relations entre les entités.
- Catégorisation des résultats.

Une recherche à 360 degrés...

Le résultat est la génération en temps réel d'une page d'accueil propre à chaque mot-clé tapé appelée « Recherche à 360° » (360° search). Pour être plus concret, détaillons ensemble la page consacrée au mot-clé « google » qui peut globalement être divisée en 4 :

- Les entités identifiées : juste au-dessous des résultats.
- Les résultats : à gauche de la page.
- Les outils de visualisation/cartographie de l'information : au milieu de la page.
- La catégorisation automatique : à droite de la page.

Les 10 entités les plus citées sont représentées ici, de Google Inc., (2 5197 citations) à Google Reader (503 citations).

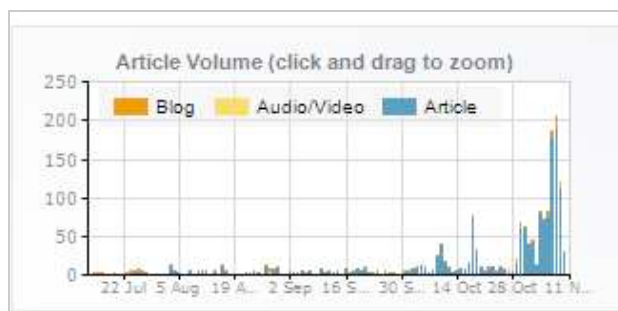


Les résultats de recherche sont présentés à l'intérieur des trois catégories principales déjà évoquées. Les actualités y sont classées dans l'ordre antéchronologique et dédoublonnées comme sur Google News. On peut voir sur cette capture d'écran les 11 articles qui sont derrière le titre « YouTubeKiller ? What YouTube Killer ». On y trouve aussi en gris les entités identifiées : *Youtube Inc., Luis Garcia, Pekka-Eric Auvinen, Google Inc.*



Les outils de visualisation sont au nombre de quatre. Lorsque vous les lancez en mode plein écran (d'un simple clic) ils se dotent de fonctionnalités complémentaires de tri et de filtrage :

- « Article volume » : présente sous la forme d'un histogramme de fréquence les volumes d'articles consacrés au mot-clé choisi (soit 1 968 pour le mot-clé « google »). Les couleurs qualifient les différents types de sources (presse, blogs, audio,'). Il est possible de zoomer sur certaines périodes en effectuant un « click and drag ». C'est le seul des quatre à ne pas proposer de mode plein écran.



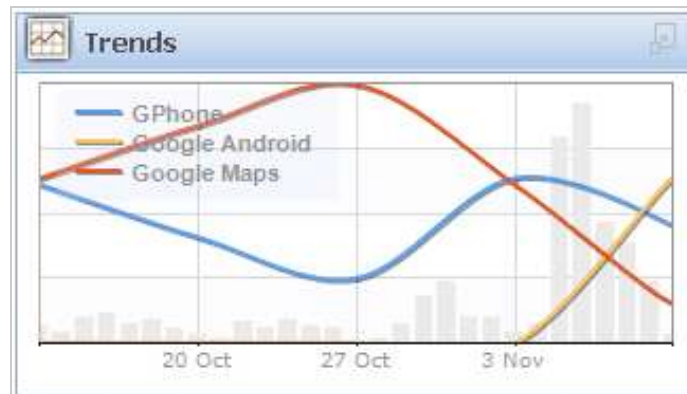
- « Network » : cartographie sous la forme d'un réseau les relations détectées entre les différentes entités extraites de l'ensemble des articles composant la page (personnalités, pays, entreprises...). En utilisant les ronds noirs on obtient les documents indiquant les relations entre deux entités. On peut augmenter ou diminuer le nombre d'entités représentées dans le réseau. Par défaut, le réseau est calculé sur une période de 4 semaines que l'on peut étendre à 6 mois ou limiter à une semaine. Il est également possible de le restreindre à une période donnée.



- « Hotspot » : représente sur une carte du monde les lieux évoqués dans les articles. La taille des points rouges est proportionnelle aux occurrences des lieux cités. En positionnant votre souris sur un de ces points le titre de l'information relative s'affiche. Là encore il est possible de zoomer dans la carte et de jouer sur la temporalité.



- « Media Attention Trends » permet de comparer les occurrences de mots-clés relatifs à votre requête principale, c'est-à-dire ce sur quoi se porte l'attention des médias. Cet outil est particulièrement intéressant car il est très configurable. Ainsi vous pouvez choisir de comparer entre eux des noms de personnes, d'organisations, d'entreprises, de villes, de pays ainsi que les mots-clés les plus cités dans les phrases-clés (voir ci-dessous). Vous pouvez également ajouter vos propres mots-clés. S'ils sont présents dans le « set » d'actualités qui a servi à créer la page ils s'afficheront dans vos courbes de tendance. Par ailleurs vous pourrez là encore jouer sur les dates.



La catégorisation automatique, ou clusterisation, se base sur les entités extraites ainsi que sur des phrases-clés et des thèmes-clés identifiés automatiquement par le système (statistiques ?).



En positionnant votre souris sur un item, vous obtenez un descriptif bref et structuré de celui-ci, par exemple la mini-biographie de Sergey Brin ou de Larry Page. Si vous cliquez sur leur nom, vous générez une page à 360° les concernant.

L'indication « Show more » permet évidemment d'afficher plus d'entités par catégories.

Si maintenant vous souhaitez affiner votre recherche et trouver dans ces 1 968 résultats ceux où la NASA est citée, il vous suffit d'entrer ce nom dans la barre de recherche et de cliquer sur « Filter » plutôt que sur « Search », vous obtenez alors 344 résultats. Vous pouvez ajouter ainsi autant de filtres que vous voulez.

Concluons avec quatre fonctionnalités qui ajoutent encore à l'expérience utilisateur :

- La recherche avancée vous permet de choisir les types de contenus dans lesquels vous voulez lancer une recherche (News, blogs, audio...), mais aussi de la limiter aux seules sources qui vous intéressent. Vous pouvez en outre la restreindre à une période donnée.
- Vous pouvez accéder directement aux outils de visualisation en les sélectionnant au-dessus de la barre de recherche avant de lancer une requête.
- En cliquant sur un article vous n'arrivez pas directement sur la page web dont il est issu mais sur un résumé automatique de l'article en question.

- Chaque requête, et donc chaque page nouvelle que vous générez et personnalisez (filtres), dispose de son propre fil RSS qui vous permet de suivre l'actualité via un agrégateur.

Pour les veilleurs et documentalistes, Silobreaker est un service d'une qualité exceptionnelle, tant au niveau de la quantité de sources intégrées qu'au niveau des fonctionnalités proposées. Son seul défaut est qu'il ne traite pour l'instant que l'anglais. Ceci mis à part il serait vraiment dommage de s'en priver, d'autant qu'il est entièrement gratuit, mais pour combien de temps ?

Christophe Deschamps

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/2007/11/novembre-2007-silobreaker-un-tonnant.html>