

<b>Domaine :</b>	<b>Recherche</b>	Référencement
<b>Niveau :</b>	<b>Pour tous</b>	Avancé

*On a beaucoup parlé de la société RecordedFuture il y a quelques mois de cela, lorsque son double financement par Google et la CIA a été évoqué dans la presse. L'ambition de cette entreprise ? Rien de moins que prédire l'avenir grâce aux informations glanées sur le Web. Nous avons eu accès à la version payante de cet outil afin de l'évaluer en profondeur. Le résultat ? Un outil prometteur même s'il s'agit avant tout d'un système d'aide à la prise de décision plutôt qu'un devin numérique en bonne et dûe forme...*

Difficile de se payer une meilleure publicité que celle qu'a assurée à Recorded Future (<http://www.recordedfuture.com/>) l'annonce de son double financement par Google et la CIA (via son fameux fond d'investissement In-Q-Tel) en juin dernier. Il faut dire que cette petite société de 16 personnes ne propose pas moins que de vous aider à prédire l'avenir via l'information de sources ouvertes qu'elle indexe et analyse en permanence. Ce type de technologie de text-mining avancé ne peut évidemment qu'intéresser les agences de renseignements puisqu'elle porte la promesse de les aider à découvrir des liens cachés entre individus, sociétés, organisations, lieux... La société ne s'y est d'ailleurs pas trompé, qui propose un blog intitulé "Intelligence Analysis" sur lequel sont décortiqués des faits d'actualité avec la technologie maison (<http://www.analysisintelligence.com/>).

Pour Google ce n'est pas beaucoup plus étonnant. Le datamining et le text mining sont en effet au cœur de son business. L'article phare de Sergey Brin et Larry Page consacré au *Pagerank* d'un Google embryonnaire en 1998 (S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine") a en effet tendance à faire oublier ceux, beaucoup plus nombreux, qu'ils ont consacré au datamining, notamment "Dynamic data mining : exploring large rule spaces by sampling", écrit un an plus tard et dans lequel ils expliquent comment, dans le but de tester une méthode d'extraction de données à partir de paniers clients sur des sites commerciaux, ils ont utilisé "des documents comme des paniers et les mots dans ces documents comme des articles"...

Chez Google l'analyse prédictive sert, entre autres choses, à anticiper vos besoins et à vous proposer des publicités en conséquence.

La version gratuite de Related Future présente en fait assez peu d'intérêt puisqu'il s'agit d'un service d'alertes par mots-clés pas très différent d'un Google Alertes. Nous avons heureusement pu obtenir un mois de test gratuit du service et en mesurer ainsi le véritable potentiel.

## ***L'interface de requêtes***

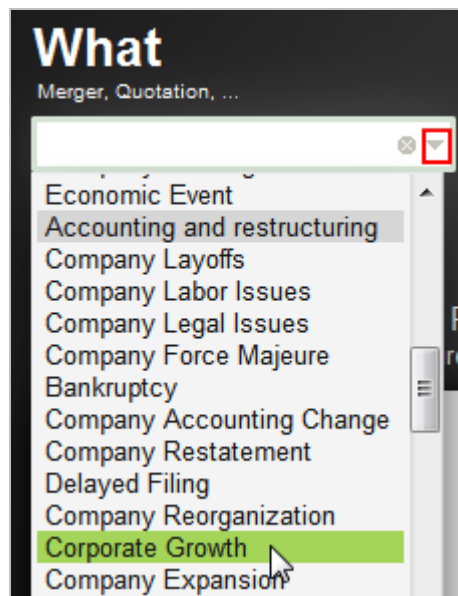
L'interface de Related Future est composée de trois boîtes de recherche thématiques.

- What ? : pour indiquer des actions et thématiques
  - Exemple : "capital market", "merger", "bankruptcy", "person travel", "movie release", "nuclear terrorism", "diplomatic relations",...
- Who?/Where? : acteur, sujet ou lieu de l'action
  - Exemple : "Obama", «New York», "China", "Nasrallah",...
- When? : pour indiquer une durée/échéance
  - Exemple : "Anytime", "last week", "next year", ...
  - Pour plus de précisions vous pouvez faire apparaître un calendrier dépliant pour y préciser vos dates.

What	Who/Where	When	
Merger, Quotation, ...	IBM, New York, Obams, ...	Now, Next week, January, ...	Clear Adv. search
Corporate Growth	Carmakers Renault SA	Anytime in the future	Display

Related Future fonctionne sur la base de crawlers qui parcourent et indexent le web en permanence et plus précisément les sites d'actualité, les blogs, les réseaux sociaux (Twitter), les sites web gouvernementaux mais aussi les bases de données financières.

La seconde étape consiste en l'extraction d'entités nommées, avec un focus sur les données temporelles, dates, durées, échéances. Lorsque vous inscrivez vos mots-clés dans le "moteur" ces mots-clés vous sont proposés automatiquement. Pour les boîtes "What ?" et "When", vous pouvez d'ailleurs partir directement d'une liste d'entités nommées en cliquant sur la petite flèche à droite.



La boîte "Who ?/Where ?" propose aussi les entités nommées "as-you-type" mais permet également une interrogation plein-texte. Ce n'est pas le cas des deux autres boîtes où vous ne pouvez qu'utiliser une entité existante sous peine de ne pouvoir lancer votre recherche. Il est en revanche possible de ne rien indiquer dans le "What ?" afin d'obtenir toutes les informations relatives à une personne/société/lieu,...

La fonctionnalité de recherche avancée vous permettra de croiser des entités nommées entre elles.

## La page de résultats

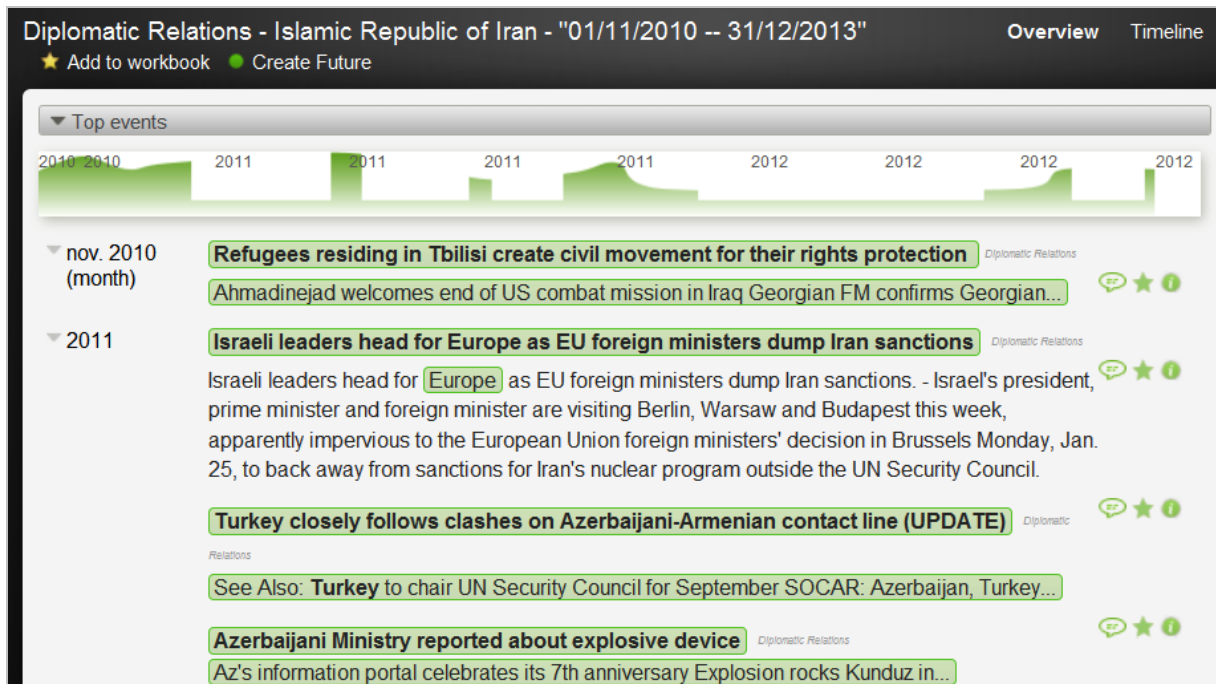
Pour les besoins de ce test nous avons choisi de lancer nos recherches sur les entités suivantes :

- What ? : Diplomatic relations
- Who ?/Where ? : Islamic Republic of Iran
- When? : 01/11/2010 – 31/12/2013

L'idée étant ici de trouver de l'information autour du nucléaire iranien.

Les résultats apparaissent sous les boîtes de recherche, dans trois types d'interfaces graphiques que nous allons maintenant détailler.

- "Overview" : interface graphique par défaut, qui se présente ainsi :



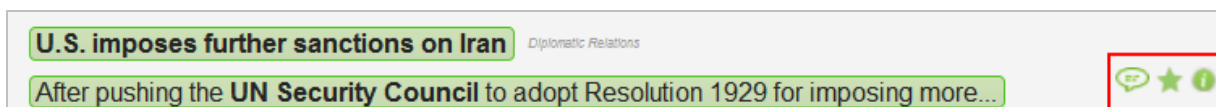
On y voit que les informations trouvées sont classées par périodes. Les blocs verts sur la frise chronologique représentent ce que les créateurs du service nomment des "Momentum", c'est-à-dire les événements/informations prenant place dans le futur, que les sources évoquent le plus au moment où vous lancez votre requête. Il s'agit donc d'une représentation de la popularité d'une information en temps-réel. Cette frise est cliquable et vous permet ainsi d'accéder directement aux informations "chaudes".

Il faut noter que la page de résultats est divisée en trois. La partie haute fait automatiquement apparaître les informations les plus en vue. Vient ensuite une partie intitulée "Top Who/Where" où vous trouverez les entités nommées les plus présentes dans les données analysées.

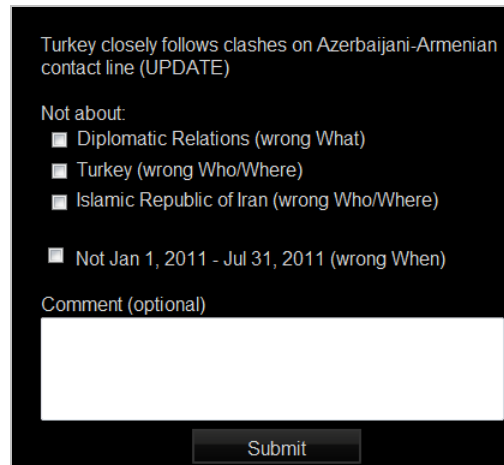


Cliquer sur l'un de ces éléments relance une nouvelle requête sur celui-ci. Enfin la troisième partie est également dotée d'une frise chronologique et permet d'accéder à la totalité des résultats.

A la droite de chaque résultat se trouvent les trois icônes suivantes :



- **Bulle** : vous permet de donner un retour sur l'information proposée (mécanisme de type crowdsourcing permettant d'améliorer les résultats au fil de l'eau).



Turkey closely follows clashes on Azerbaijani-Armenian contact line (UPDATE)

Not about:

- Diplomatic Relations (wrong What)
- Turkey (wrong Who/Where)
- Islamic Republic of Iran (wrong Who/Where)
- Not Jan 1, 2011 - Jul 31, 2011 (wrong When)

Comment (optional)

Submit

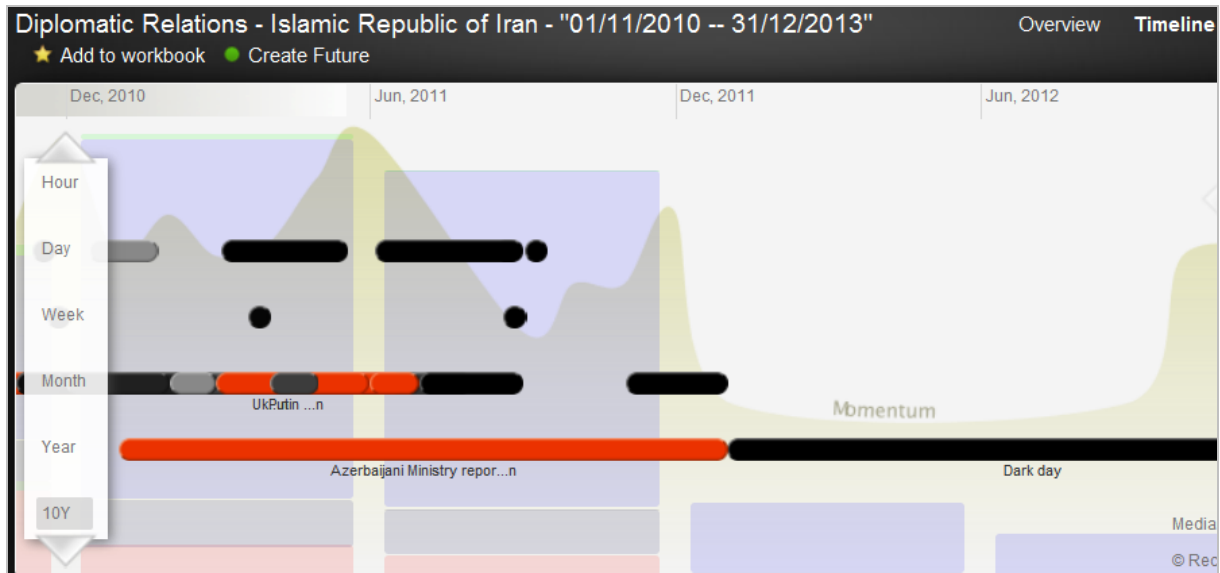
- **Etoile** : ajoute l'élément sélectionné au "Workbook". Il s'agit d'un panier virtuel qui vous permet d'enregistrer les articles ou les recherches les plus intéressants pour y revenir ultérieurement.



- **i** : affiche des informations complémentaires sur l'article sélectionné.

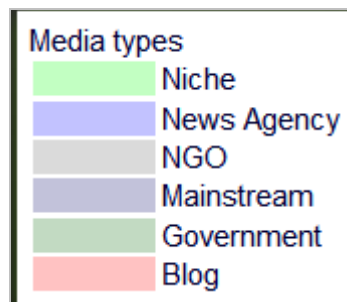


**Représentation graphique "Timeline"** : comme son nom l'indique, cette représentation est également basée sur une frise chronologique mais permet d'explorer les résultats plus fine.

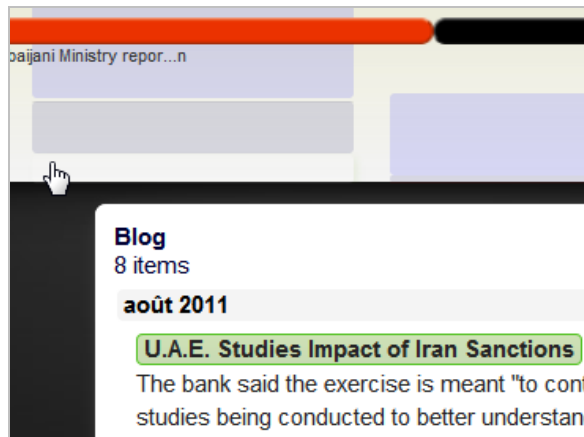


C'est d'abord la possibilité présente à gauche de l'interface, d'afficher ces données sur différentes périodes, heure, mois, année,... Cela peut permettre de replacer une information dans son contexte ou, au contraire, de focaliser sur une période donnée.

Les codes couleurs des blocs rectangulaires en arrière-plan correspondent aux grands types de sources dont sont issus les résultats.



Cliquer sur l'un de ces rectangles permet de filtrer les résultats par type de sources.

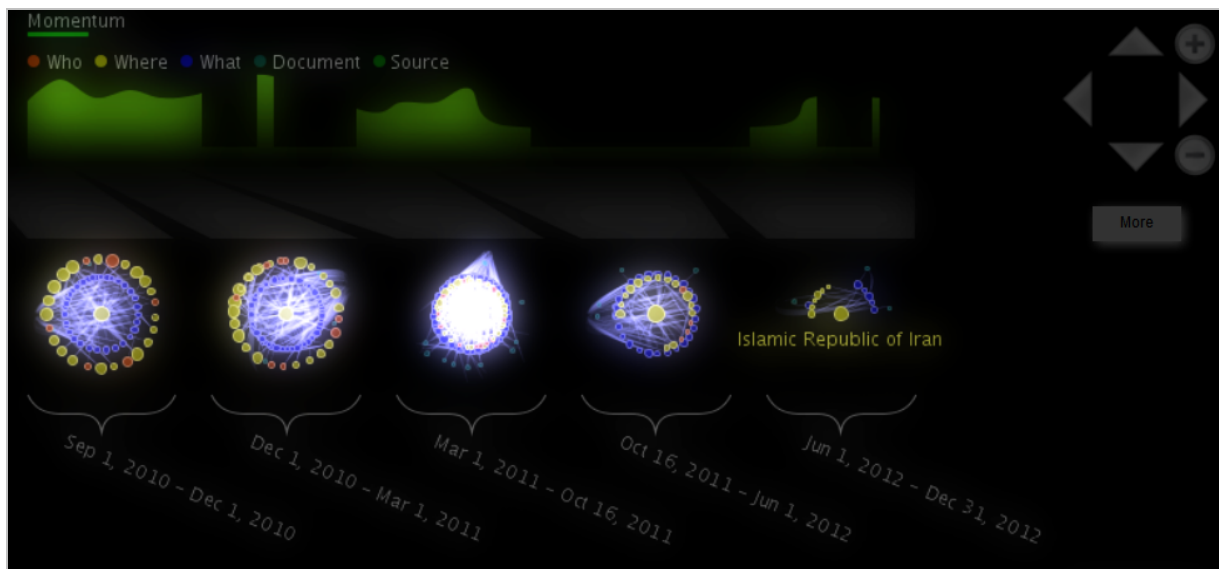


Les points uniques représentent un évènement lié à une date spécifique alors qu'une "barre" indique qu'il s'étend sur une durée plus longue. En cliquant sur ces points ou barres on fait apparaître une fenêtre dans laquelle les articles qui les constituent s'afficheront. Related Future fait aussi de l'analyse de sentiments. Ainsi le gris représente le neutre, le rouge-orangé le négatif et le noir le positif.



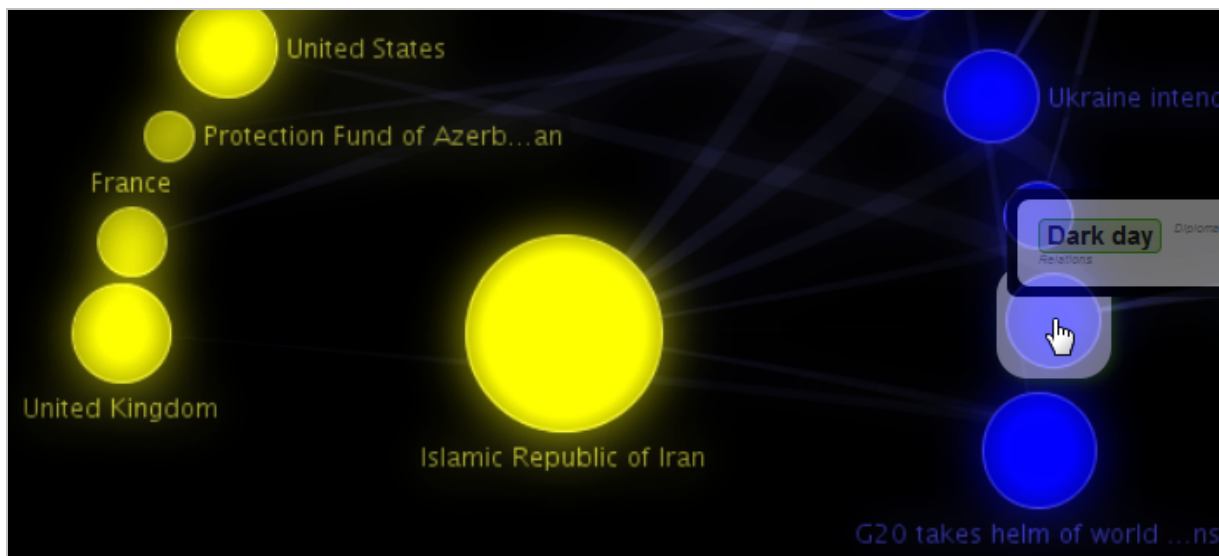
A noter que la timeline peut être "avancée" ou "reulée" par simple "glisser-déposer".

**Représentation "Network"** : cette représentation graphique positionne votre requête "à l'intérieur" de son réseau relationnel d'entités nommées. Ainsi chaque entité évoquée dans les sources prises en compte sera ici représentée graphiquement et dans une période donnée.



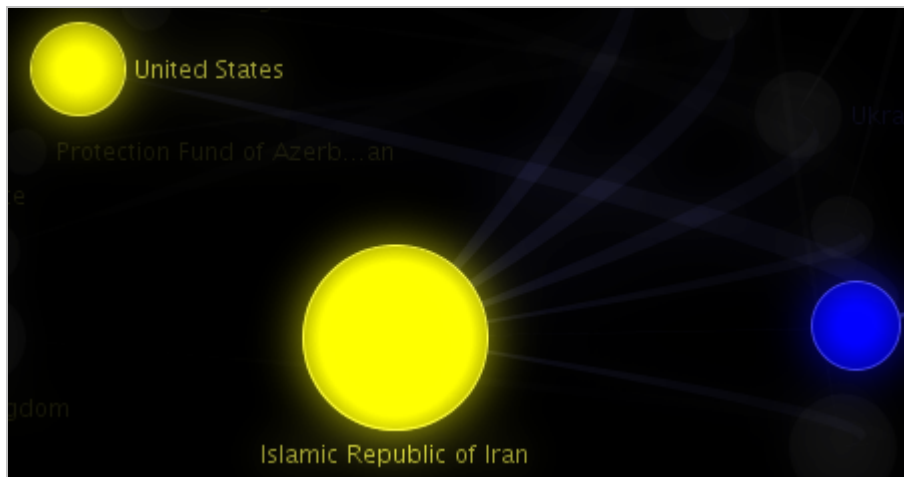
Des codes couleurs (voir en haut de l'image ci-dessus) permettent de savoir à quel type d'entité on a affaire. Afin de pouvoir exploiter ces représentations il faut zoomer sur celles qui vous intéressent grâce aux symboles + et - présents en haut à droite de la fenêtre. Les flèches vous servant bien sûr à vous déplacer dans la carte.

Pour que la lecture ne soit pas trop complexe nous avons choisi de zoomer sur le réseau le plus à droite.



On constate que les titres des cercles deviennent visibles et que l'on peut cliquer dessus pour avoir accès aux articles qui les composent.

En fonction de leur couleur, les cercles ne disposeront pas des mêmes fonctionnalités. Ainsi si vous cliquez sur un rond jaune et choisissez "Near" vous aurez, par exemple, la possibilité d'afficher les éléments du réseau avec lesquels il est en relation.



Ici "United States" est en relation avec "Islamic Republic of Iran" mais aussi avec "Dark Day" (cercle bleu), c'est-à-dire le 11 septembre. En cliquant sur ce dernier on arrive à un article de BBC News qui cite effectivement les trois entités.

### Eight years since the 'dark day'



Dealing with the legacy of the attacks is one of Barack Obama's toughest challenges

By Roger Hardy  
BBC Middle East and Islamic affairs analyst

**Across the United States, the memory of 9/11 - that fateful Tuesday in 2001 when suicide attacks struck New York and Washington - is still very much alive.**

This year the annual commemoration of the attacks in the US has been given a new name - the National Day of Service and

On a donc effectué une requête croisée en cliquant sur des entités nommées représentées par des objets graphiques.

Enfin, le bouton "More" situé en dessous des flèches permet de filtrer les éléments de la carte pour ne faire apparaître que :

- les sources ;
- les documents ;
- What ? ;
- Who ?/Where ?

Pour finir ce descriptif fonctionnel notez que :



- Vous pouvez à tout moment transformer votre page en une alerte en cliquant sur le bouton "Create Future".



- Vous pouvez transformer vos résultats en un fichier KML afin d'afficher les entités nommées géographiquement (sur Google Maps par exemple).
- Vous pouvez embarquer une représentation graphique afin de la rendre accessible sur une page web. Ces deux fonctionnalités sont présentes en bas à gauche de chaque page de résultats.

## Quid des résultats ?

Comme on le voit, les fonctionnalités de Related Future sont nombreuses et bien pensées. Reste à savoir si les résultats proposés sont pertinents...

Les tests que nous avons menés sur l'Iran sont moyennement convaincants. On voit en effet que certains articles apparaissent dans les résultats pour de mauvaises raisons. Ainsi nous avons eu plusieurs articles provenant du journal en ligne azerbaïdjanais en anglais Trend (<http://en.trend.az>) qui ne citaient pas l'Iran. En parcourant la page on s'aperçoit toutefois que le mot est bien cité en tant que tête de rubrique.

**TREND** | Company news | Par | Expert Council | Pro

Now in Baku: 19:51 (GMT+05:00), Tuesday

Home News Economics Viewpoint

Advanced Search

Azərbaycan | Русский | العربية | فارسی | Azerbaijan South Caucasus **Iran** Central Asia

< Back to home page / Azerbaijan / Foreign policy

### Refugees residing in Tbilisi create civil movement for

01.09.2010 18:06

Georgia, Tbilisi, Sept.1 / Trend, N. Kirtskhalia /

Refugees created their own organization - Civil Movement for protection of their rights. To this end, more than 100 people, who lost their homes in Abkhazia and South Ossetia gathered near the

Il y a de fortes chances pour que les algorithmes n'aient pas fait la différence. Nous l'avons également testé sur un thème plus "populaire" (Products – Google – 01/10/2010-31-12-2013) et les résultats nous ont semblé plus précis.

Par ailleurs il faut s'entendre sur le terme "prédire l'avenir". Il n'est d'ailleurs pas employé par les créateurs du service qui le qualifient de "*temporal analytics engine*". S'il n'est évidemment pas la boule de cristal que tout le monde attend (en vain), sa structuration par période permet en revanche d'avoir, par exemple, une bonne vision de ce que prépare la concurrence dans tel ou tel pays (implantation, joint-venture, nomination,...).

Notons enfin qu'il ne travaille pour l'instant qu'à partir de sources anglophones.

Au final Related Future est un service qui nous semble intéressant. Il souffre bien sûr des faiblesses évoquées ci-dessus, ainsi que d'une lenteur parfois exaspérante (que l'on peut sans doute imputer à sa relative jeunesse). Il offre cependant une véritable interface d'aide à l'analyse de l'information, car c'est de cela dont il s'agit et de rien d'autre. L'outil informatique, quel qu'il soit, ne fait pas d'analyse au sens ou un humain peut en faire (et ne le fera jamais), il fait du traitement automatique du langage et c'est déjà beaucoup. Ici une interface graphique bien pensée peut apporter une véritable valeur ajoutée au travail d'un analyste à condition qu'il accepte d'en payer le prix... C'est un autre point bloquant que nous voyons à ce service. Il faut en effet chiffrer la valeur ajoutée qu'il est susceptible de vous apporter et vous assurer qu'elle correspond bien à la somme, déjà conséquente, d'environ 1800 \$/an (149\$/mois). Comparer notamment cette valeur à celle que des solutions alternatives, mêmes basiques, peuvent vous apporter. Chaque contexte étant particulier il se peut que l'outil soit utile dans un cas et pas dans l'autre. La bonne nouvelle est qu'il est proposé en mode SaaS et que sa tarification est mensuelle, avec la possibilité de stopper l'abonnement à n'importe quel moment. Un test est donc aisé à mettre en œuvre et ce sera, selon nous, la seule manière de voir s'il correspond vraiment à vos besoins.

**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://blog-abonnes.abondance.com/2010/09/recordedfuture-la-boule-de-cristal-du.html>