

# Ces start-ups qui investissent le champ de l'évaluation des sources et contenus

Carole TISSERAND-BARTHOLE

*Comme nous avons pu le voir précédemment, certains géants du Web et les outils de recherche et de veille professionnels commencent progressivement à intégrer des fonctionnalités permettant d'évaluer la fiabilité des sources, notamment en mettant en place des partenariats avec des start-ups qui se positionnent sur l'évaluation des sources et des contenus.*

*Nous avons interviewé trois acteurs sur le marché français se positionnant sur ce créneau, avec des approches intéressantes mais divergentes. Nous avons par ailleurs identifié les autres acteurs en place, que nous analyserons dans cet article.*

**TrustedOut :**

l'IA au service des corpus de sources

*TrustedOut « exploite les dernières avancées technologiques de machine learning, de web crawling, et de micro-services pour pallier les limites et biais de la curation de contenu manuelle, et fournir des corpus de sources constamment à jour et dignes de confiance. ». TrustedOut a récemment signé un partenariat avec Digimind, qui propose en option l'intégration du produit dans sa plateforme Digimind Intelligence.*

*La start-up a été créée en 2018 par Freddy Mini, ex-CEO de Netvibes, que nous avons interviewé.*

**NETSOURCES : D'où vous est venue l'idée de TrustedOut ?**

**Freddy MINI :**

Le nom : *TrustedOut* vient de "if it's not Trusted In, it can not be Trusted Out" (si vous n'avez pas confiance dans ce qui entre, vous ne pouvez pas avoir confiance dans ce qui sort). Les Américains disent "Garbage in, garbage out".

**Notre vision :** la perte de confiance, actuelle et brutale, dans l'intégrité de l'information a des conséquences directes sur l'éducation, les opinions et, plus important encore, la prise de décision. Comme nous pensons que le vrai et le faux sont subjectifs et souvent, personnels, le *fact-checking* magique ne redonnera pas une confiance universelle. Le *profiling* des auteurs est également illusoire (en plus d'être perturbant) car les auteurs ont plusieurs (ou aucune) identité, qu'ils ne sont pas experts en tout et que leur expertise est temporaire et là encore subjective.

La seule, et la bonne solution, est donc de profiler le *publisher* (éditeur), et en particulier les valeurs de son logo, de sa marque. Car le logo est l'épine dorsale, l'élément vital et donc il protège, défend à tout prix le *Publisher*. En profilant les valeurs du *Publisher*, vous pouvez définir votre confiance en lui.

Et comme la confiance est personnelle (personne ne peut vous forcer à avoir confiance en quelque chose), il fallait fournir une base de données des profils des sources d'information pour que les analystes, professionnels du marketing, PR Managers puissent définir leurs corpus d'information (l'ensemble du contenu qui sera utilisé).

**Executive Manager**  
Business Analyst,  
Product Marketer,  
Communication or  
PR Manager at  
**ACME**

**For ACME's product launch, I need publications I trust defined by:**

- Financial & Business Services industry, specialized in Advertising,
- Focused on B2B,
- In the USA, UK and France
- That are not politically, nor religiously, nor humorously / satirically oriented,
- In business for more than 2 years.
- Can be a government companies but no blogs nor wikis.

**TrustedOut provides, live, ACME with corresponding media, sources and article abstracts to:**

- Feed Social Intelligence
- Secure Brand Safety
- Optimize PR efforts

**Incoming:**  
Project query

**Outgoing:**  
media, sources, abstracts

**Professional advertising market in US, UK and France**  
Corpus live for 51 days (4/12/19) | deactivate | billing | duplicate | pulse | delete

217 media | 1378 sources | 34k article abstracts | 879K since 1/7/19

**Edit the definition of your corpus**  
Add or remove condition to refine what you trust

Basic mode | Expert mode

Taxonomic rank	Classification	Advertising	Dedicated
Marketing target	is	Business to Business	
Countries (2 paid options)	are	US, UK, France	
Political orientation	is	None	
Religious orientation	is	None	
In business for	More than	2 years	
Company organization	is not	Self-Publisher	

Back Save

**Live Taxonomy**  
Top category

Corpus analytics

Business  
Lifestyle  
Society  
Tech  
Politics  
Economy  
Marketing  
Finance  
Entrepreneurship  
Therapie

Figure 1. Interface de TrustedOut

## INTERVIEW

Alors qu'il y a un fort intérêt pour les algorithmes d'intelligence sémantique (BI) et la programmation publicitaire, rien n'existait dans l'intelligence des corpus. D'où notre slogan : *Corpus Intelligence*.

**NETSOURCES :** Pourriez-vous nous expliquer en quelques lignes le fonctionnement de l'outil et quelle pourrait être son utilité pour des professionnels de l'information ?

**Freddy MINI :**

Vous créez un corpus par une requête en ajoutant des conditions (pays, taxonomie, typologie des sources, etc.) – (voir figure 1),

*TrustedOut* vous montre à chaque étape le nombre de media, sources, nombre d'articles par jour en moyenne et nombre d'abstracts que nous avons en archive pour cette requête. Une fois satisfait, vous pouvez soit télécharger ces media et sources ou connecter votre système pour que *TrustedOut* l'alimente automatiquement et en permanence.

**NETSOURCES :** Sur votre site, vous définissez *TrustedOut* comme une « media database ». Qu'est-ce que le terme media englobe exactement? Et de quelles informations précises dispose-t-on pour chaque média ?

**Freddy MINI :**

Nous couvrons, pour l'instant, tout ce qui a un statut de *Publisher*, donc news, magazines, revues... quelle que soit la configuration technique du *Publisher*.

Rien ne nous empêche de profiler des sources payantes ou autres. Il nous faut simplement un accès aux sources pour les classer.

Les informations par source sont à la fois intangibles (*data collection*: nom, propriétaire, CA, traffic...) et tangibles (*content classification*: taxonomie, perçu comme toxique (*fake news, junk science...*)). En tout plus de 60 champs et 400 catégories.

**NETSOURCES :** L'outil se focalise pour l'instant sur les US et la France. D'autres pays vont-ils être ajoutés prochainement et si oui, dans quel délai ?

**Freddy MINI :**

UK, Espagne, Mexique, Portugal, Brésil et aussi Canada

francophone et anglophone d'ici la fin de l'année

**NETSOURCES :** Vous disposez déjà d'un partenariat avec *Digimind*. Avez-vous d'autres partenariats avec d'autres plateformes de veille ? *TrustedOut* est-il également proposé comme un outil à part sans lien avec des plateformes ?

**Freddy MINI :**

Nous travaillons actuellement à des partenariats avec les agences publicitaires et *trading desks* pour leur fournir des listes blanches qui résolvent le problème aigu de la *Brand Safety*.

*TrustedOut* sera également ouvert en libre-service en novembre avec un tarif mensuel de 500 euros pour 1 corpus et 300 euros pour chaque autre corpus.

Storyzy : du fact-checking à l'évaluation des sources, il n'y a qu'un pas

*Storyzy catalogue les sites et vidéos de désinformation. Nous sommes allés interviewer Pierre Albert Ruquier, Directeur Marketing et co-fondateur de Storyzy*

**NETSOURCES :** Comment est né *Storyzy* et comment a-t-il évolué depuis son lancement ?

**Pierre Albert RUQUIER :**

A l'origine *Storyzy* s'appelait *Trooclick*, fondé par des professionnels de l'information et entrepreneurs. J'ai par exemple été journaliste durant 25 ans (presse écrite, *BFM Business*, co-fondateur de *6médiadigital*, *France-Soir*, *Sipa News*). Avec *Trooclick* nous avons, grâce au traitement automatique des langues et au *machine learning*, commencé à automatiser la vérification des informations (*fact-checking*). Nous avons même sorti un *add-on* qui s'appelait *Glitch Spotter* et qui donnait le taux de fiabilité d'un article donné dans le domaine des informations économiques.

Les raisons pour lesquelles nous avons abandonné cette piste sont multiples : trop complexe techniquement d'étendre le système à d'autres thématiques que le *business*, pas de possibilité de développer un *add-on* sur les mobiles, un modèle économique difficile (personne prêt à payer et volumes insuffisants pour monétiser par la publicité), peu d'erreurs factuelles dans les articles sur l'économie.