



HAL
open science

Retour d'expérience : un outil professionnel de veille pour une grande institution de la République

Mathieu Andro, Marie-Cécile Haenel, Caroline Pount-Biset, Agnès Frère, Olivier
Le Cocguen, Stéphanie Mezbourian, Djamila Chamekh, Brigitte Berquier

► **To cite this version:**

Mathieu Andro, Marie-Cécile Haenel, Caroline Pount-Biset, Agnès Frère, Olivier Le Cocguen, et al.. Retour d'expérience : un outil professionnel de veille pour une grande institution de la République. 2020. <hal-02488615>

HAL Id: hal-02488615

<https://hal.science/hal-02488615v1>

Preprint submitted on 23 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Retour d'expérience : un outil professionnel de veille pour une grande institution de la République

Mathieu ANDRO, documentaliste

dans une grande institution de la République

mathieuandro@yahoo.fr

Marie-Cécile HAENEL, documentaliste

dans une grande institution de la République

Caroline POUNT-BISET, documentaliste

dans une grande institution de la République

Agnès FRERE, documentaliste

dans une grande institution de la République

Olivier LE COCGUEN, documentaliste

dans une grande administration de la République

Stéphanie MEZBOURIAN, documentaliste

dans une grande institution de la République

Djamila CHAMEKH, documentaliste

dans une grande institution de la République

Brigitte BERQUIER, documentaliste

dans une grande institution de la République

Introduction et contexte

L'institution dans laquelle l'outil professionnel de veille KB Crawl a été expérimenté n'est pas nommée dans le présent article pour des raisons d'indépendance et de déontologie. Il s'agit d'une institution de près de 2000 personnes pour laquelle travaillent une cinquantaine de documentaliste en régions et à Paris.

La veille, sur des thématiques très variées, y occupe la majeure partie des services documentaires et cette partie est croissante. Elle y est menée de manière artisanale et non homogène à partir d'outils gratuits de surveillance (agrégateurs de flux RSS, alertes par courriels ou encore surveillance non automatisée). Elle donne lieu à des livrables de veille divers (CMS, Sharepoints, PDF, courriels). Les paramètres et les résultats des veilles ne sont pas partagés et la consommation qui est faite de ces veilles est partiellement voire totalement inconnue selon les cas.

Dans ces conditions, un outil professionnel de veille a été expérimenté par les 8 documentalistes signataires du présent article afin de :

- disposer d'un outil commun et partagé de veille collaborative
- surveiller des sites non compatibles avec la technologie RSS d'autant que cette technologie est absente sur un nombre croissant de sources¹
- surveiller des parties de sites
- compléter l'utilisation de bases bibliographiques de dépouillement de la presse dans l'usage veille qui en est fait car leurs couvertures ne sont pas exhaustives en termes de sources ni de contenu et car le dépouillement est différé par rapport au dépouillement d'un crawler qui est immédiat
- challenger le recours à des prestations extérieures par nos services de communication concernant les retombées presse (veille image)
- **partager les sources et les informations** au sein du réseau de documentalistes et éviter que plusieurs documentalistes surveillent les mêmes sources et valident les mêmes informations, renforcer l'esprit collaboratif entre les documentalistes. Éviter les doublons et redondances entre les veilles
- **permettre une navigation intuitive pour les usagers** en leur offrant plusieurs angles d'accès aux différentes veilles, au-delà de leurs seuls périmètres.
- harmoniser la diffusion des veilles sur des supports plus fonctionnels et réduire le temps actuellement consacré à éditorialiser
- **proposer une offre de nouveaux services** de manière proactive par l'automatisation de la sélection de certaines veilles pour lesquelles nous n'avons pas reçu de commande directe mais dont les besoins sont récurrents
- **capitaliser** de manière pérenne des informations détectées
- analyser des corpus de veille (bibliométrie, text mining, datavisualisation)
- obtenir des statistiques de consultation des veilles par les usagers
- rendre plus efficace les dispositifs de veille (rendement et précision)

¹ <https://bibliotheques.wordpress.com/2011/10/11/vers-la-fin-du-rss>

Méthode

Les solutions Coexel, DigiMind et KB Crawl ont été identifiées. Il a été décidé d'expérimenter KB Crawl avec 8 licences sur l'année 2019 et de mesurer les résultats obtenus.

Résultats de l'expérimentation

Sur près de 300 sources paramétrées, il a été difficile d'en paramétrer une quinzaine en raison de leurs structures ou car les éditeurs de ces sources sont parvenus à identifier le robot venu les visiter.

Au total, il y a eu 10 286 articles publiés sur la plateforme en 2019. Et 1708 articles ont été envoyés via 144 Newsletters. Ces contenus ont généré 276 connexions d'utilisateurs en 2019 (dont 38 connexions via smartphones qui ont consulté 3728 articles sur la plateforme, dont 243 à partir d'un smartphone et 349 à partir d'une newsletter envoyée par courriel. Si on met en relation avec les 10 286 articles publiés en 2019, on peut considérer que 36 % des articles ont été lus au moins une fois. C'est un résultat encourageant, d'autant qu'il est ascendant dans le temps.

Enquêtes de satisfaction

Enquête auprès des veilleurs ayant expérimenté KB Crawl

Voici comment les fonctionnalités proposées par KB Crawl ont été appréciées par les 5 veilleurs qui ont expérimenté l'outil de veille et qui se sont exprimés :

| Fonctionnalité | Appréciation moyenne |
|--|--|
| Surveiller des sites non compatibles avec la technologie RSS Surveiller des sites accessibles via des logins Filtrer les contenus par mots clés Obtenir des statistiques de consultation détaillées des veilles par les usagers | Très utile (note moyenne de 3,8 / 4) |
| Permettre à l'utilisateur de partager ou de valider directement des articles Publier des articles originaux non crawlés Permettre l'accès sur smartphone Capitaliser les corpus de veille Surveiller des sites et des parties de sites | Très utile (note moyenne de 3,6 / 4) |
| Automatiser la mise en forme et l'envoi des newsletters Partager les sources au sein du réseau | Utile ou très utile (note moyenne de 3,5 / 4) |
| Partager les informations collectées au sein du réseau Eviter les redondances entre veilles | Utile (note moyenne de 3,4 / 4) |
| Harmoniser la diffusion des veilles Produire des analyses textuelles | Utile (note moyenne de 3,2 / 4) |
| Produire des tags automatiquement à partir de vocabulaires et de thesaurii | Utile (note moyenne de 3 / 4) |

| | |
|--|---------------------------|
| Challenger les bases de dépouillement de la presse | Utile |
| Challenger la veille externalisée sur les retombées presse | (note moyenne de 2,8 / 4) |
| Challenger les envois manuels de sommaires de revues | Utile |
| | (note moyenne de 2,6 / 4) |

Les expérimentateurs préfèrent plutôt ce type d'outil à un agrégateur de presse (note moyenne de 1,8 / 4 où 1 représente une préférence pour ce type d'outil et 4 une préférence pour un agrégateur de presse)

Globalement, ils considèrent, comme les usagers, qu'il est judicieux de se doter de ce type d'outil (note de 3,2/ 4 où 4 présente le plus haut niveau de volonté de se doter de ce type d'outil).

Voici les raisons invoquées :

| Commentaires |
|---|
| Gagner en efficacité afin d'alléger la charge de travail et d'accroître nos services quantitativement (plus grande couverture) et qualitativement (production d'analyses) |
| Parce qu'il est d'utilisation très souple (une fois que l'on a bien compris et assimilé toutes les possibilités de paramétrages qui sont quand même assez complexes), très convivial, qu'il permet un travail collaboratif très efficace et qu'il permet de gagner beaucoup de temps dans le rubriquage ainsi que la mise en forme des documents et newsletters pour un rendu très ergonomique et agréable à l'oeil, permettant la création de cartes géographiques et de graphiques de datavisualisation. Il permet également d'effectuer des recherches avec un moteur très efficace, basé sur les capacités de l'outil qui s'avèrent très fines en matière de catégorisation et de taggage des documents. Enfin il permet d'avoir plus facilement des retours des utilisateurs (avec les commentaires) ainsi que des statistiques d'usages très pertinentes. |
| Kb Crawl ne peut remplacer un agrégateur de presse pour les demandes de dossiers documentaires ou de recherches ponctuelles. Qui nécessitent l'accès aux archives. A moins d'y faire des dossiers permanents sur tous les sujets susceptibles d'être programmés. Je perçois la puissance de l'outil mais je n'ai pas encore pu en tester toutes les fonctionnalités. |
| <ul style="list-style-type: none"> - Facilité d'accès à l'information pour les utilisateurs - Gain de temps phénoménal à l'éditorialisation des newsletters de veille - Donner l'info à des groupes restreints d'utilisateurs - Possibilité pour les utilisateurs de collaborer à la collecte des informations - Automatisation des tâches répétitives et sans plus value documentaires - Possibilité de tenir à jour des statistiques (sur les sources les plus parlantes, sur les mots clés les plus consultés, sur les visites des utilisateurs. Ces statistiques nous permettent de réajuster nos contenus. - Présence d'un moteur de recherche efficace - Possibilité de donner une visibilité aux veilles des régions qui ne trouveraient pas d'espace sur l'Intranet |

Enquête auprès des usagers ayant consommé des veilles sous KB Crawl

27 utilisateurs ont répondu à notre enquête. Les utilisateurs interrogés sont globalement satisfaits de l'accès à l'information sur la plateforme (note moyenne de 2,8 / 4. 4 étant la note la plus satisfaisante) et surtout de son aspect (note moyenne de 3,11 / 4). L'intérêt d'une lecture des veilles sur smartphone est un peu moins exprimé par les utilisateurs mais reste néanmoins et en moyenne satisfaisant (2, 23 / 4).

L'angle thématique (2,07 / 4) est celui qui est préféré par les utilisateurs, suivi de l'angle territorial (1,85), de l'angle métier (1,81), de l'angle organisationnel (1,69), puis de l'angle de la communauté de travail (1,56).

De manière générale, « pour consulter les informations sur un portail », « ce type de plateforme » est largement préféré à l'Intranet, à Sharepoint ou à tout « autre type de portail (ex Scoop it, Netvibes, ...) » et « un envoi email via ce type d'outil de veille » est largement préféré à « un envoi email via l'Intranet », à « un envoi de email avec pièce jointe » ou à « une alerte Sharepoint » :

Et pour conclure, de manière globale, les 27 utilisateurs interrogés considèrent à 3, 7 / 4 en moyenne qu'il est « profitable de se doter de ce type d'outil » pour les raisons suivantes :

| Commentaires |
|--|
| <i>« Cet outil me semble correspondre aux nouveaux usages des personnels. Pratique (envoi par Email) intuitif et dans les codes actuels (usage de visuels).Il me semble indispensable d'adapter la forme aux usages, cet outil permet de produire notamment un panorama de presse plus agréable à lire et susciter de l'intérêt. »</i> |
| <i>« Automatisation de certains segments du processus de veille qui génère des gains de temps et permet de se recentrer sur de l'apport de valeur ajoutée ou l'élargissement de l'offre de services. Visions partagée sur l'ensemble des veilles disponibles Possibilité de partager facilement des articles ou des informations »</i> |
| <i>« notamment pour effectuer des veilles sur des sujets auxquels l'Intranet offre peu ou pas de place (problématiques régionales, par ex) »</i> |
| <i>« Plus simple et centralise tout »</i> |
| <i>« Mutualisation des veilles documentaires et des dossiers documentaires des régions »</i> |
| <i>« Cet outil est censé faciliter la veille qui doit rester pertinente et présenter l'information de façon conviviale. Il est nécessaire que cet outil donne envie de consulter les informations sur la plateforme »</i> |
| <i>« Fait gagner du temps. On balaye l'information disponible rapidement »</i> |
| <i>« La newsletter nous informe de façon essentielle même si les articles ne sont pas tous ouverts, on sait qu'on les a vu et que l'on pourra les retrouver sur la plateforme »</i> |
| <i>« Le rubriquage des informations est un plus »</i> |
| <i>« L'aspect du mail est parfait »</i> |
| <i>« L'outil est meilleur que les outils précédents »</i> |

Seuls 2 utilisateurs interrogés n'en sont pas convaincus :

| Commentaire |
|--------------------|
|--------------------|

« Pitié !!! trop d'infos tuent l'info ALORS pensez : trop de sources et de manip TUENT L'ENVIE de se connecter !!! pourquoi l'INTRANET avec un acces au COIN DES PROS dérange ? JE NE VAIS JAMAIS SUR CETTE PLATEFORME je ne sais même pas où sont les codes mentionnés ... J'ai assez de sources à me souvenir, de codes d'accès, de pratiques et/ou localisations des messages des fichiers joints.... FAISON SIMPLE : une source : l'INTRANET et un COIN DES PROS où échanger ET CONSERVER. Un Exemple : les statistiques que 3 services et personnes nous font parvenir !! »

Résultats par rapports aux objectifs initiaux

Avec KB Crawl, la mise en surveillance d'une source est la partie la plus délicate et chronophage. Sur la source la plus « simple » qui a été testée, il faut au moins 5 minutes pour paramétrer une source simple :

| | |
|---------------------|---------------------|
| 1er crawl | 39 secondes |
| MacroRecorder | 32 secondes |
| 1er scraper | 49 secondes |
| 2e crawl | 42 secondes |
| 2e scraper | 49 secondes |
| autres paramétrages | 20 secondes |
| Total | 231 secondes |

Mais parfois il faut **beaucoup plus de temps** pour paramétrer une source de façon pertinente car la technicité est souvent très différente d'une source à l'autre. Au cours de l'expérimentation, 150 sources ont été mises en surveillances (près de 10 heures de travail) soit la quasi-totalité de celles que nous avons indiquées dans notre cahier des charges.

KB Crawl nous permet de challenger l'utilisation de bases bibliographiques de dépouillement de la presse dans l'usage veille (et non recherche) qui en est fait car leurs couvertures ne sont pas exhaustives en termes de sources, car le dépouillement est non immédiat, et car elles sont relativement coûteuses

De manière générale, nous pouvons supposer que la consommation de ces bases sera maintenue pour les besoins en recherches documentaires sur des années anciennes et pour le dépouillement de sources auxquelles nous ne sommes pas abonnés. En revanche, cette consommation des bases de dépouillement de la presse pourrait diminuer au moins dans l'usage qui en est fait pour les veilles.

KB Crawl pourrait aussi permettre de challenger le recours à des prestations extérieures par nos services de communication concernant les retombées presse (veille image). Il est notamment beaucoup plus réactif pour certaines sources qu'il peut repérer et signaler immédiatement

KB Crawl a permis d'harmoniser la diffusion des veilles sur des supports plus fonctionnels et réduire le temps actuellement consacré à éditorialiser. Ainsi, la durée d'éditorialisation d'une newsletter est bien moindre avec KB Crawl, sans aucune mesure par rapport aux autres outils testés jusque-là (INTRANET, Adobe Acrobat ou NuancePdf, email simple, page htm, présentation ppt, etc...)

Fonctions ou fonctionnalités non comparées car non remplies en l'absence d'un outil de veille

- partager ses sources et ses informations au sein du réseau de documentalistes pour éviter que plusieurs documentalistes surveillent les mêmes sources et valident les mêmes informations, renforcer l'esprit collaboratif entre les documentalistes et disposer d'un support de listes de veilles en cours accessible de façon intuitive
- éviter les doublons et redondances entre les veilles
- harmoniser la diffusion des veilles sur des supports fonctionnels pour réduire le temps consacré à éditorialiser
- proposer une offre de service de manière proactive en automatisant
- capitaliser de manière pérenne des informations détectées
- analyser des corpus de veille (bibliométrie, text mining)
- obtenir des statistiques de consultation des veilles par les usagers
- surveiller des sites non compatibles avec la technologie RSS
- surveiller des parties de sites

Autres fonctionnalités testées

Les fonctionnalités suivantes ont été testées avec succès :

- Accès réservé de certains articles à certaines personnes ou groupes de personnes
- Publication automatique d'articles (retombées médias, presse et Twitter)
- Envoi de newsletters entièrement automatisée, semi-automatisée, manuelle
- Analyses textuelles (text mining) à partir du vocabulaire géographique déjà présent sur la plateforme puis à partir du thésaurus et de la liste des administrations : taggage automatique et datavisualisation sur corpus

Conclusions

Bilan : intelligence artificielle, text mining, crowdsourcing

L'outil est très innovant et nous a permis d'expérimenter des technologies proches de l'**intelligence artificielle** :

- Automatisation de la veille en amont (surveillance) et en aval (diffusion)
- Apprentissage progressif, pour le robot, de ce qu'il doit filtrer afin de :
 - réduire le bruit des informations non pertinentes
 - sélectionner les zones précises à copier

- Possibilité d'envisager l'automatisation de certaines veilles temporairement en cas d'absence d'un veilleur, sous réserve de tester plus avant afin d'éviter les problèmes importants de bruit généré dans les tests

Nous avons également expérimenté des technologies de **text mining** :

- Tags automatiques à partir de vocabulaires (géographie, administrations, thésaurii...)
- Production d'analyses textuelles sous forme de datavisualisation

Enfin, nous avons aussi la possibilité de recourir à une forme de **crowdsourcing**, c'est-à-dire d'externalisation d'une partie de l'activité directement auprès de l'utilisateur :

- permettre la collaboration d'utilisateurs des informations collectées
- publication d'informations originales par les utilisateurs et renseignement humain

Difficultés rencontrées

L'investissement dans un outil professionnel de veille est un investissement financier mais encore d'avantage humain important.

Offre de services sur la plateforme (pistes de développement)

Niveau 1 : l'utilisateur souhaite simplement pouvoir republier des articles lui-même pour les faire connaître à ses collègues ou encore publier des articles originaux

Niveau 2 : l'utilisateur souhaite qu'un documentaliste mette en place une mise en surveillance de sources éventuellement filtrées par des mots clés et que les articles ainsi collectés soient systématiquement et automatiquement publiés

Niveau 2.1 : sans aucune curation

Niveau 2.2 : avec une curation a posteriori et un apprentissage du robot

Niveau 3 : l'utilisateur souhaite se charger lui-même de sélectionner, parmi les articles collectés, ceux qu'il publiera sur la plateforme

Niveau 4 : l'utilisateur souhaite qu'un documentaliste se charge d'effectuer cette curation a priori

Niveau 5 : l'utilisateur souhaite obtenir une analyse de corpus

Niveau 5.1 : analyse textuelle par datavisualisation

Niveau 5.2 analyse intellectuelle des informations collectées sous la forme d'une synthèse périodique produite par des documentalistes

Evolution prévisible des activités

Un accompagnement au changement sera nécessaire afin de passer à terme, par exemple, de tâches chronophages d'éditorialisation à des tâches de maintenance de paramétrage des sources ou de mise en forme des maquettes de Newsletters. Le gain en temps pourra également être utilisé afin d'étendre d'avantage notre couverture en veilles.

Certaines activités pourraient périlcliter :

- GED sur des revues *en ligne* qui pourraient désormais être crawlées
- Dépouillements de certaines revues dans un SIGB
- Surveillance « manuelle » des sources
- Copiés-collés « manuels » des articles
- Mise en forme (éditorialisation) « manuelle » des Newsletters
- Envois « manuels » des Newsletters

D'autres, au contraire, devraient se développer :

- Utilisation de KB Crawl afin de **surveiller automatiquement le bon accès aux abonnements** et anticiper ainsi sur les réclamations des usagers en communiquant en amont sur notre Intranet en cas de problème d'accès
- Mises en formes des **maquettes** de Newsletters
- **Design** de la plateforme de diffusion
- Paramétrage de la **mise en surveillance de nouvelles sources** dans KB Crawl
- **Data visualisations** et **cartographies**
- **Maintenance** et amélioration itérative du paramétrage des **sources** déjà surveillées
- Production d'analyses textuelles de corpus de veille
- Rédaction d'articles originaux d'analyse ou de synthèse

Concernant la veille, 4 types de fonctions pourraient se développer, permettant de faire évoluer les postes de veilleurs qui continueraient à assurer toute la chaîne de veille :

- **Animation, direction de projet** (responsable du pôle services au public et veilles ?)
 - piloter le développement du projet au-delà de l'expérimentation
 - former les documentalistes et les usagers
 - contrôler que les documentalistes n'aient pas trop de documents en attente
 - maintenir le crawler et la plateforme de veilles
 - modérer les publications
 - arbitrer le plan de tags
 - assurer la relation avec l'éditeur
 - alimenter un tableau de bord d'indicateurs
 - etc...
- **Sourcing** (mettre en place le dispositif de veille)
- **Curation** (valide les informations collectées)

- **Design** (améliore et personnalise la mise en forme)

Afin d'accompagner ce changement, une formation complète pourra être assurée dans le cadre du marché. Par la suite, les collègues les plus expérimentés pourront accompagner ceux qui le sont moins.

Gains pour les documentalistes

- **Diminuer la charge actuelle de travail** et par automatisation de la surveillance et de la mise en forme.
- **gagner du temps avec le partage facilité** des documentalistes mais aussi avec **l'implication des usagers dans les sélections** grâce au modèle collaboratif (fonctionnalités de RSE, réseau social d'entreprise)
- **Suivre les statistiques** avec la possibilité de mesurer la consommation des veilles de manière détaillée
- Gagner en reconnaissance en offrant un service de meilleure qualité
- **Proposer de nouveaux services** et de **nouveaux produits** non existants jusqu'ici, notamment des services à plus haute **valeur ajoutée** (analyses intellectuelles et analyses textuelles, data visualisation, text mining) plus gratifiants qui permettront un meilleur développement personnel, un développement de compétences, et éventuellement aussi une évolution de carrière
- Monter en compétence dans le cœur de métier en étant consommateur des veilles partagées

Gains pour les usagers

- avec les fonctionnalités de RSE (réseau social d'entreprise) :
 - Partager eux-mêmes directement des informations du web
 - Produire eux même directement leurs propres informations
 - Sélectionner eux-mêmes directement des informations extraites par le dispositif de veille
- Gain en qualité : confort de lecture (ergonomie et smartphone), l'utilisateur connaît souvent mieux le sujet que le veilleur et sa sélection est plus judicieuse
- Gain en couverture plus large des sujets grâce au temps gagné pour les veilleurs
- Ouverture sur les autres veilles

Perspectives

L'expérimentation nous semble concluante. Les statistiques de consultation, les enquêtes auprès des utilisateurs et des veilleurs nous incitent à déployer l'outil. Néanmoins, dans la mesure où l'outil ne fait pas l'unanimité parmi les documentalistes, il est proposé de demander aux documentalistes qui le souhaitent de se manifester pour obtenir une licence et une formation. Les documentalistes qui souhaitent poursuivre avec d'autres outils seront libres de le faire. Concernant l'accès en lecture, il sera déployé pour l'ensemble de l'institution.