

h e g

Evaluation de la qualité des résultats de la veille : comparaison des plateformes Digimind, Sindup et MyTwip

Mémoire de recherche réalisé par :

Fanny QUIQUEREZ

Floriane MINANA

Sous la direction de :

Hélène MADINIER, Professeure HES

Genève, le 17 janvier 2018

Master en Sciences de l'information

Haute École de Gestion de Genève (HEG-GE)

Déclaration

Ce mémoire de recherche est réalisé dans le cadre du Master en Sciences de l'information de la Haute école de gestion de Genève. L'étudiant accepte, le cas échéant, la clause de confidentialité. L'utilisation des conclusions et recommandations formulées dans ce travail, sans préjuger de leur valeur, n'engage ni la responsabilité de l'auteur/des auteurs, ni celle de l'encadrant.

« Nous attestons avoir réalisé le présent travail sans avoir utilisé des sources autres que celles citées dans la bibliographie. »

Fait à Genève, le 17 janvier 2018

Fanny Quiquerez



Floriane Minana



Remerciements

Nous souhaitons adresser nos chaleureux remerciements aux personnes qui nous ont aidées à mener à bien ce travail :

Hélène Madinier, sans qui ce projet n'aurait pas existé, que nous souhaitons remercier pour sa confiance et son expertise.

Maurizio Velletri et Julien Gobeill pour leur aide et leurs précieux conseils durant certaines phases du projet.

Stéphanie Minana, sans qui l'analyse des données récoltées aurait été impossible. Nous la remercions infiniment pour son aide, son enthousiasme et ses compétences expertes sur Excel et Query.

Un grand merci à nos proches qui nous ont soutenues et accompagnées, de près ou de loin, durant l'élaboration de notre projet.

Résumé

Les plateformes de veille sont des « outil[s] développé[s] par des sociétés spécialisées qui [ont] pour but de collecter, d'analyser, synthétiser et diffuser l'information que l'on peut trouver sur Internet » (Winello 2011). Notre projet de recherche consiste à évaluer la qualité de la veille des plateformes Digimind, Sindup et MyTwip sur la base des résultats retournés. Il est considéré comme innovant car s'il existe de nombreuses publications qui comparent ces outils selon les fonctionnalités à disposition, aucune étude ne propose à l'heure actuelle d'effectuer un classement selon notre approche. Cette recherche s'inscrit dans le cadre d'un mandat proposé par la Haute école de Gestion afin de contribuer au développement d'une plateforme de veille.

Notre étude doit répondre aux deux objectifs suivants :

- Proposer une méthode innovante d'évaluation de la qualité des plateformes de veille Digimind, Sindup et MyTwip sur la base des résultats retournés,
- Définir quelle plateforme est la plus performante au regard de cette évaluation.

Pour ce faire, elle se compose de trois grandes parties :

Une première partie théorique, composée de la revue de littérature, permet de cerner le cadre conceptuel, aborder les notions et métriques propres à l'évaluation de la recherche d'information, prendre connaissance des trois plateformes étudiées et procéder à un état de l'art des comparatifs d'outils de veille existants.

Une deuxième partie technique expose la méthodologie que nous avons développée afin de mener à bien ce comparatif. Nous y détaillons nos plans de veille et de recherche, les indicateurs sélectionnés pour définir la qualité des plateformes, leurs paramétrages, ainsi que la récolte des résultats.

Enfin, en dernière partie, nous analysons les résultats de la veille obtenus sur chacune des plateformes et nous les interprétons de manière à définir la plateforme la plus performante.

Conscientes des limites de notre méthodologie et des résultats obtenus, nous proposons en conclusion un certain nombre de recommandations pour de prochains projets.

Mots-clefs :

Veille, Plateforme de veille, Comparatif, Méthode d'évaluation de la qualité, Analyse des résultats.

Table des matières

| | |
|---|------------|
| Déclaration | i |
| Remerciements | ii |
| Résumé | iii |
| Table des matières | iv |
| Liste des tableaux | vii |
| Liste des figures | vii |
| 1. Introduction | 1 |
| 1.1 Contexte | 1 |
| 1.2 Nature du projet | 1 |
| 1.3 Problématique | 2 |
| 1.4 Perspectives | 3 |
| 2. Revue de la littérature | 4 |
| 2.1 Cadre conceptuel | 4 |
| 2.1.1 La veille..... | 4 |
| 2.1.2 Le cycle de la veille..... | 5 |
| 2.1.3 La recherche d'information | 6 |
| 2.1.4 Les outils de veille | 6 |
| 2.1.5 Les plateformes de veille | 9 |
| 2.1.6 La notion de qualité | 9 |
| 2.1.7 La notion d'évaluation | 9 |
| 2.1.8 La notion de source | 10 |
| 2.1.9 La notion d'indicateurs | 10 |
| 2.1.10 La notion de résultats | 10 |
| 2.2 Les outils de veille | 10 |
| 2.2.1 Les plateformes de veille évaluées..... | 10 |
| 2.2.2 Les comparatifs d'outils de veille | 13 |
| 2.3 L'évaluation de la recherche d'information | 18 |
| 2.3.1 Recherche d'information et système de recherche d'information | 18 |
| 2.3.2 Évaluation d'un SRI | 19 |
| 2.3.3 Les métriques de la RI | 19 |
| 3. Méthodologie de la recherche | 21 |
| 3.1 Approche générale | 21 |
| 3.2 Le ciblage | 21 |
| 3.2.1 Analyse des besoins et recherches exploratoires | 21 |
| 3.2.2 Plans de veille..... | 22 |
| 3.2.3 Pré-test | 23 |

| | | |
|------------|---|-----------|
| 3.3 | Sourcing et plans de veille..... | 23 |
| 3.3.1 | Le choix des sources | 23 |
| 3.3.2 | Plans de recherche | 24 |
| 3.4 | Paramétrage des plateformes..... | 25 |
| 3.5 | Collecte des données | 26 |
| 3.5.1 | Structure des dossiers de veille au sein des plateformes..... | 26 |
| 3.5.2 | Phase de veille | 27 |
| 3.5.3 | Choix des indicateurs | 27 |
| 3.5.4 | Saisie des résultats..... | 28 |
| 3.5.5 | Echantillonnage | 29 |
| 3.5.6 | Traitement des données | 31 |
| 3.6 | Analyse des résultats | 33 |
| 4. | Présentation et analyse des résultats | 34 |
| 4.1 | Volumétrie..... | 34 |
| 4.1.1 | Volumétrie des sources manuelles (ou SM) | 34 |
| 4.1.2 | Volumétrie des sources automatiques (ou SA) | 36 |
| 4.2 | Pertinence..... | 37 |
| 4.2.1 | Pertinence des sources manuelles (ou SM) | 38 |
| 4.2.2 | Pertinence des sources automatiques (ou SA) | 38 |
| 4.3 | Précision | 40 |
| 4.3.1 | La précision des sources manuelles (ou SM)..... | 40 |
| 4.3.2 | La précision des sources automatiques (ou SA) | 41 |
| 4.4 | Complémentarité..... | 42 |
| 4.4.1 | La complémentarité des sources manuelles (ou SM)..... | 42 |
| 4.4.2 | La complémentarité des sources automatiques (ou SA) | 43 |
| 4.5 | Rapidité de capture..... | 44 |
| 4.5.1 | La rapidité de capture des sources manuelles (ou SM) | 45 |
| 4.5.2 | La rapidité de capture des sources automatiques (ou SA)..... | 45 |
| 4.6 | Tableaux récapitulatifs des performances des plateformes | 47 |
| 4.6.1 | Récapitulatif pour les sources manuelles (ou SM) | 47 |
| 4.6.2 | Récapitulatif pour les sources automatiques (ou SA)..... | 48 |
| 5. | Conclusion | 49 |
| 5.1 | Réponses à nos questions de recherche | 49 |
| 5.1.1 | La méthode d'évaluation employée | 49 |
| 5.1.2 | La performance des plateformes | 49 |
| 5.2 | Limites de la méthode et des résultats obtenus | 49 |
| 5.2.1 | Le paramétrage des plateformes | 49 |
| 5.2.2 | Le choix des indicateurs | 50 |

| | | |
|-------------------|---|------------|
| 5.2.3 | Autres éléments de notre méthode..... | 50 |
| 5.2.4 | Les résultats | 51 |
| 5.3 | Proposition d'autres méthodes d'évaluation | 51 |
| 6. | Bibliographie..... | 53 |
| Annexe 1 : | Sujets de veille | 58 |
| Annexe 2 : | Plans de veille | 60 |
| Annexe 3 : | Plans de recherche | 64 |
| Annexe 4 : | Cartes mentales | 95 |
| Annexe 5 : | Equations de recherche | 99 |
| Annexe 6 : | Présentation des résultats par sujet de veille | 126 |

Liste des tableaux

| | |
|--|----|
| Tableau 1 : Tableau de contingence des documents..... | 20 |
| Tableau 2 : Les indicateurs définis pour évaluer la qualité des plateformes | 27 |
| Tableau 3 : Exemple de tableur de saisie des résultats | 29 |
| Tableau 4 : Méthode d'échantillonnage par sélection aléatoire par ordinateur | 29 |
| Tableau 5 : Calcul automatisé de la précision..... | 30 |
| Tableau 6 : Evolution de la pertinence des résultats..... | 31 |
| Tableau 7 : Croisement des résultats pertinents entre les 3 plateformes | 32 |
| Tableau 8 : Croisement des dates de collecte entre les 3 plateformes..... | 32 |
| Tableau 9 : Récapitulatif des résultats par plateforme et sujet de veille | 32 |
| Tableau 10 : Tableau récapitulatif pour les sources manuelles | 47 |
| Tableau 11 : Tableau récapitulatif pour les sources automatiques | 48 |

Liste des figures

| | |
|---|----|
| Figure 1 : Les composantes de l'intelligence économique | 4 |
| Figure 2 : Le cycle de la veille | 6 |
| Figure 3 : Le trèfle fonctionnel des outils de veille | 7 |
| Figure 4 : Le modèle RI | 18 |
| Figure 5 : SM : Nombre et type de résultats par plateforme | 34 |
| Figure 6 : SM : Nombre et type de résultats par plateforme et par sujet de veille | 35 |
| Figure 7 : SA : Nombre et type de résultats par plateforme | 36 |
| Figure 8 : SA : Nombre et type de résultats par plateforme et par sujet de veille | 37 |
| Figure 9 : SM : Nombre de résultats pertinents par plateforme | 38 |
| Figure 10 : SM : Nombre de résultats pertinents par plateforme et par sujet de veille | 38 |
| Figure 11 : SA : Nombre de résultats pertinents par plateforme | 39 |
| Figure 12 : SA : Nombre de résultats par plateforme et par sujet de veille | 39 |
| Figure 13 : SM : Précision par plateforme | 40 |
| Figure 14 : SM : Précision par sujet et par plateforme | 40 |
| Figure 15 : SA : Précision par plateforme | 41 |
| Figure 16 : SA : Précision par sujet et par plateforme..... | 41 |
| Figure 17 : SM : Complémentarité des résultats retournés | 42 |
| Figure 18 : SM : Complémentarité des résultats retournés par sujets | 43 |
| Figure 19 : SA : Complémentarité des résultats retournés | 43 |
| Figure 20 : SA : Complémentarité des résultats retournés par sujet..... | 44 |
| Figure 21 : SM : Rapidité de capture Digimind VS Sindup..... | 45 |
| Figure 22 : SA : Rapidité de capture MyTwip VS Sindup | 45 |
| Figure 23 : SA : Rapidité de capture Digimind VS Sindup | 46 |
| Figure 24 : SA : Rapidité de capture Digimind VS MyTwip | 46 |
| Figure 25 : SA : Rapidité de capture Digimind VS MyTwip VS Sindup | 47 |

1. Introduction

1.1 Contexte

Face au nombre toujours plus croissant d'informations en ligne, les outils de veille sont devenus incontournables pour surveiller les pages web ainsi que filtrer et collecter des données liées à un sujet précis. De typologie et périmètre fonctionnel différents, la palette d'outils de veille est très vaste, allant de simples alertes e-mail jusqu'à des logiciels payants contrôlant un large éventail de sources. Ces derniers, appelés également plateformes de veille, sont beaucoup utilisés par les entreprises qui souhaitent surveiller le secteur dans lequel elles opèrent. En effet, "la veille [leur] permet de saisir des opportunités de développement et de détecter des menaces potentielles, et plus globalement d'accroître [leur] compétitivité" (UFC et al. 2017).

En 2013, la Société i+1 et la Haute école de gestion de Genève ont mis en place une plateforme de veille webSO/Inélio (Innovarc [2016]). Mené conjointement par la filière Information documentaire de la Haute école de gestion de Genève (HEG GE), le Centre Tesnière de l'Université de Bourgogne-Franche-Comté (UBFC) et l'entreprise Erdil à Besançon, ces différents acteurs travaillent actuellement sur le projet webSO+, qui a pour objectif de développer de nouvelles fonctionnalités pour cette plateforme de veille. Ce projet vise à "mettre à disposition des entreprises un outil de veille performant [...], qui permet[te] de collecter, traiter et diffuser l'information nécessaire à la surveillance du marché" (UFC et al. 2017).

La plateforme webSO+ possédera comme fonctionnalités supplémentaires un traitement automatique des informations utiles issues de la veille ainsi qu'un suivi de l'e-reputation. Par ailleurs, des innovations seront mises en place grâce à l'utilisation des techniques de traitement automatique des langues. L'ensemble de ces fonctionnalités sera adapté, au niveau de la terminologie, aux domaines d'activités des entreprises concernées par le projet (UFC et al. 2017).

1.2 Nature du projet

Notre projet s'inscrit dans le contexte exposé précédemment et propose de comparer trois plateformes de veille : Digimind, Sindup et MyTwip, logiciels concurrents de la plateforme webSO/Inélio, auxquels la Haute école de gestion est abonnée. Notre analyse a pour objectif d'évaluer la qualité de ces trois outils par l'intermédiaire des résultats retournés.

Nous baserons notre évaluation sur trois sujets de veille (cf. Annexe 1) définis par Hélène Madinier et Arnaud Gaudinat, enseignants à la Haute école de gestion et directement impliqués dans le projet webSO+. Ces trois sujets ont été choisis car ils touchent à des thématiques bien distinctes, font appel à des sources variées et correspondent à des types de veille fréquemment effectués (veille autour d'une entreprise, veille des tendances dans un domaine et veille sur un marché émergent). Par ailleurs, ces sujets intéressent directement certains mandants et nos résultats seront pour eux sources d'information.

1) Le premier sujet concerne Lemsys, société qui conçoit et fabrique des appareils de test pour semi-conducteurs de puissance, située à Plan-les-Ouates. Il s'agira d'effectuer une veille à propos de ses plus importants concurrents, des investissements et des fusions-acquisitions des acteurs-clé du marché ainsi que de l'évolution du marché des semi-conducteurs de puissance aux USA.

2) Le deuxième sujet traite de l'évolution des métiers en Sciences de l'information en Suisse et dans les pays voisins. La veille devra particulièrement se focaliser sur les offres d'emploi, l'identification des attentes et le suivi de l'évolution des besoins.

3) Enfin, le dernier sujet porte sur l'IoT (Internet of things) ou l'industrie 4.0. L'objectif sera d'obtenir une idée plus précise de l'industrie 4.0 adaptée au domaine de la santé et de l'industrie pharmaceutique (fournisseurs, clients, pratiques), essentiellement en Suisse.

Aucune étude comparable n'ayant déjà été réalisée et les résultats étant destinés à être utilisés dans le cadre du projet de plateforme webSO+, notre recherche est de type exploratoire appliquée.

Elle comportera trois volets :

- Théorique : Nous ferons une revue de littérature afin de bien cerner le cadre conceptuel de notre étude et de prendre connaissance des trois plateformes que nous comparerons. Puis nous ferons un état de l'art des écrits comparant les outils de veille.
- Technique : Nous paramètrons les trois plateformes et nous formulerons les requêtes propres à chaque outil. Puis, nous effectuerons une veille de quelques mois afin d'obtenir suffisamment de données.
- Analytique : Nous analyserons les résultats de la veille obtenus sur chacune des plateformes à l'aide d'indicateurs que nous aurons préalablement définis, puis nous les interpréterons afin de définir la plateforme la plus performante.

1.3 Problématique

Notre recherche vise les objectifs généraux suivants :

- Proposer une méthode innovante d'évaluation de la qualité des plateformes de veille Digimind, Sindup et MyTwip sur la base des résultats retournés
- Définir quelle plateforme est la plus performante au regard de cette évaluation

Nos questions de recherche se traduisent donc ainsi :

- Quelle méthode et quels indicateurs permettent d'évaluer la qualité des plateformes de veille Digimind, Sindup et MyTwip sur la base des résultats retournés ?
- Quelle est la plateforme la plus performante au regard de cette évaluation ?

1.4 Perspectives

Les éditeurs de plateformes de veille telles que celles étudiées dans ce projet sont très nombreux sur le marché et la compétition entre eux est importante. Ainsi, les résultats de notre étude seront analysés et comparés à la future plateforme webSO+ par les parties prenantes du projet, et s'il apparaissait que ce nouvel outil est plus performant, cela offrirait un argument de taille pour sa commercialisation.

2. Revue de la littérature

2.1 Cadre conceptuel

Notre problématique faisant appel à un certain nombre de concepts, le cadre conceptuel ci-après vise à définir et présenter les notions clés de notre étude.

2.1.1 La veille

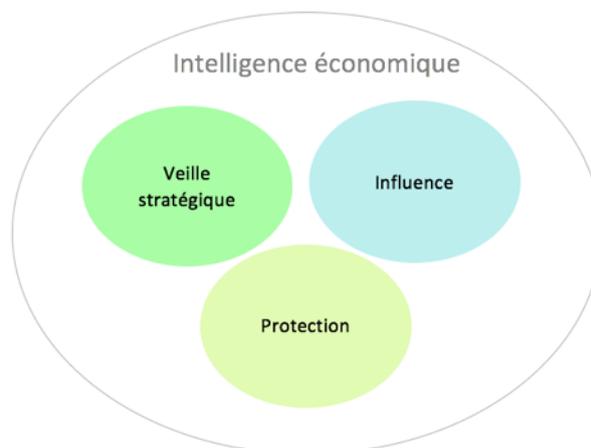
Le concept de veille stratégique s'inscrit dans le champ plus large d'intelligence économique définit comme étant

« l'ensemble des actions coordonnées de recherche, de traitement et de distribution en vue de son exploitation, de l'information utile aux acteurs économiques. Ces diverses actions sont menées légalement avec toutes les garanties de protection nécessaires à la préservation du patrimoine de l'entreprise, dans les meilleures conditions de qualité, de délais et de coût. » (Commissariat Général du Plan 1994, p. 11).

L'intelligence économique représente un processus anticipatif ayant pour finalité l'aide à la prise de décisions et au choix de l'orientation stratégique de l'entreprise (Balmisse, Meingan 2008). Elle comporte 3 facettes (Hassen 2014) :

- La veille stratégique : collecte d'informations
- L'influence : utilisation offensive des informations
- La protection : mesures de protection (juridiques, informatiques, opérationnelles...) des informations pour garantir l'intégrité, la disponibilité et la confidentialité.

Figure 1 : Les composantes de l'intelligence économique



La veille stratégique, communément appelée veille, représente donc l'une des composantes clé de l'intelligence économique en la nourrissant d'informations (Fosso 2012).

Il s'agit d'un processus de collecte, traitement et diffusion d'informations (Association des professionnels de l'information et de la documentation 2015) dont le but est de permettre à l'entreprise ou l'organisation de détecter, analyser et suivre les évolutions ou menaces liées à son environnement et de pouvoir y réagir (Association française de normalisation 1998). Grâce à sa valeur stratégique, la veille doit fournir des données de manière régulière et personnalisée à un destinataire bien précis (Université européenne de Bretagne 2016).

La veille n'est pas une tâche ponctuelle mais une « activité continue en grande partie itérative visant à une surveillance active de l'environnement technologique, commercial, etc., pour en anticiper les évolutions » (Association française de normalisation 1998, p. 6). Il s'agit donc d'un processus dynamique n'ayant de sens que dans la durée (Université européenne de Bretagne 2016).

La veille émerge dans les années 1980 (Fosso 2012) et s'est progressivement intégrée aux activités de nombreuses entreprises et organisations quel que soit leur taille ou le secteur d'activité (Thomas 2008). L'une des raisons de l'importance toujours croissante de cette activité est due à la mondialisation qui a amené les entreprises à être confrontées à toujours davantage de concurrence ainsi qu'à la surcharge informationnelle liées aux technologies web (Madinier 2017).

La notion de veille stratégique implique la surveillance d'un environnement dans sa globalité. Il existe néanmoins des déclinaisons répondant à des enjeux stratégiques très spécifiques. Ainsi par exemple, la veille dite sociétale s'intéressera aux comportements humains, la veille juridique aux normes et règlements et la veille image, à la perception de l'organisation par les clients et fournisseurs (Balmisse, Meingan 2008).

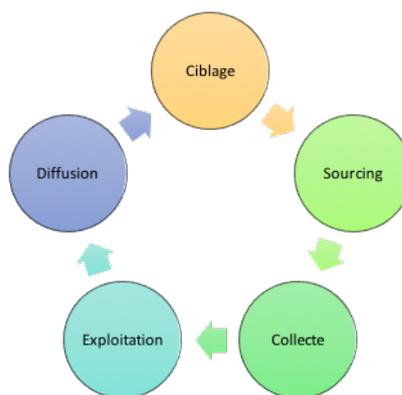
2.1.2 Le cycle de la veille

Étant une activité continue et itérative, le processus de veille est représenté sous forme d'un cycle, appelé cycle de veille ou cycle du renseignement (Madinier 2016). Il existe un grand nombre de modèles proposés dans la littérature professionnelle dont le nombre et la dénomination des étapes peuvent diverger.

On retrouve néanmoins toujours les mêmes phases principales, à savoir :

- Ciblage : Définition précise du ou des sujets de la veille et de sa finalité (Association française de normalisation 1998), du périmètre d'investigation et des besoins (Université européenne de Bretagne 2016). C'est notamment lors de cette étape qu'est élaboré le plan de veille qui permet de formaliser les besoins identifiés (Madinier 2016).
- Sourcing : Identification et choix des sources d'informations les plus appropriées pour répondre aux besoins identifiés (Dupin 2014). Lors de cette étape, il s'agit de définir le plan de recherche (Madinier 2016) et les différentes équations de recherche si nécessaire.
- Collecte : Récolte systématique des informations (Dupin 2014).
- Exploitation : Traitement et analyse des informations collectées (Université européenne de Bretagne 2016).
- Diffusion : Transmission des informations aux personnes concernées par le biais d'un produit de veille et d'un support de diffusion (Dupin 2014 ; Université européenne de Bretagne 2016).

Figure 2 : Le cycle de la veille



Dans le cadre de notre étude, nous intervenons dans les phases de *ciblage*, de *sourcing* et de *collecte*, et, notre analyse portant sur les résultats retournés par le dispositif de veille, nos mesures porteront sur la phase *exploitation*.

2.1.3 La recherche d'information

Dans la littérature professionnelle, on assiste régulièrement à une confusion entre les notions de recherche d'information et de veille.

Les deux notions se rejoignent par le fait que leur activité est de rassembler des informations pertinentes pour répondre au besoin informationnel d'un utilisateur. Néanmoins, leur finalité est tout à fait différente. Alors que la recherche d'information a pour but de répondre ponctuellement à une question précise d'un usager (Cacaly, Le Coadic 2004), la veille, elle, est un processus sur le long terme qui vise à identifier les informations signalant une évolution dans un environnement bien précis (Maniscalco 2016).

La recherche d'information fait cependant partie intégrante du processus de veille, et notamment lors de l'utilisation de plateforme. Elle intervient dans la phase de *sourcing* puisque c'est lors de cette étape que sont choisies les sources d'information et que les différentes requêtes sont élaborées.

Notre méthodologie s'appuyant sur certains concepts inhérents à la recherche d'information et à son évaluation, nous approfondirons cette notion dans notre revue de littérature.

2.1.4 Les outils de veille

Un outil de veille est un logiciel qui permet la mise en pratique du processus de veille. Dans la littérature, la notion d'*outil de veille* est la plus souvent utilisée, mais on retrouve également l'expression *logiciel de veille* qui va dans le même sens (Dualé 2010).

L'offre disponible sur le marché est abondante et regroupe des outils très différents les uns des autres proposant des prestations très larges (Balmisse, Meingan 2008 ; Dualé 2010). Afin d'avoir un aperçu global du marché, Balmisse et Meingan (2008) proposent de le segmenter selon trois dimensions : le périmètre fonctionnel, le modèle économique et le type d'architecture.

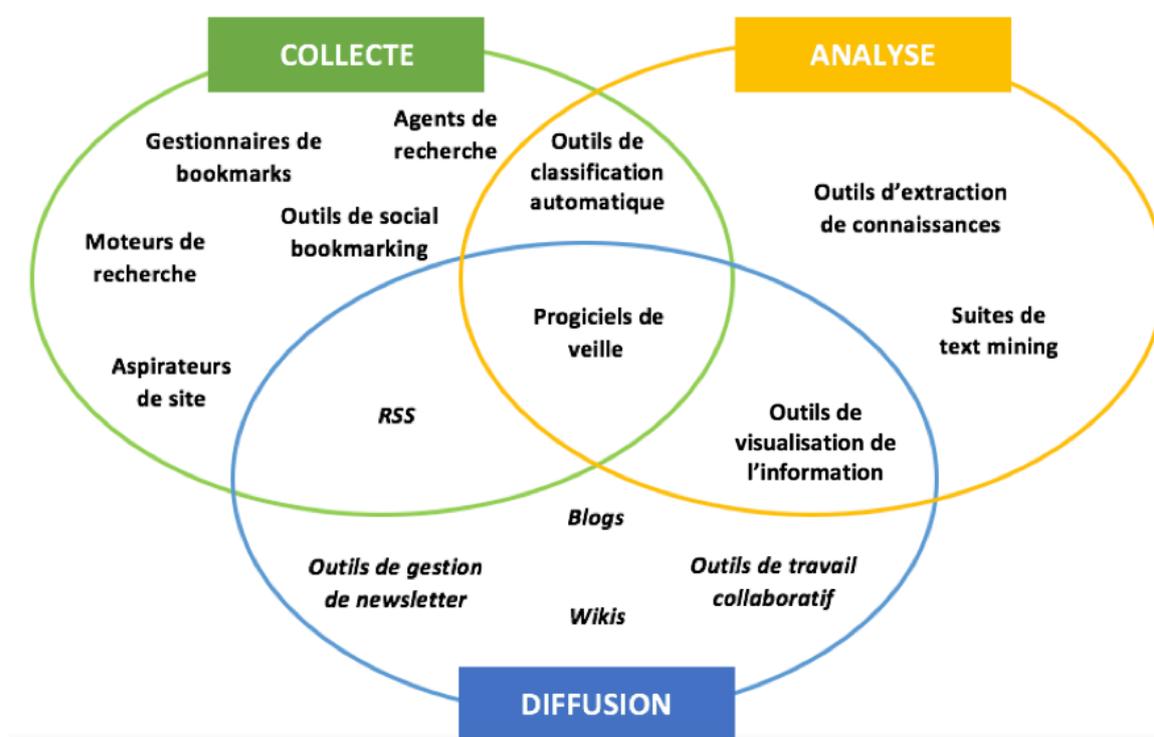
2.1.4.1 Le périmètre fonctionnel des outils de veille

Les fonctionnalités propres aux outils de veille peuvent être classées en trois catégories, la collecte, l'analyse et la diffusion, qui peuvent être associées de différentes manières (Dualé 2010).

- Collecte : regroupe toutes activités visant à identifier, recueillir et gérer les informations de manière systématique.
- Analyse : concerne les activités de validation des informations, d'identification des contenus intéressants, de réalisation de rapports de synthèse...
- Diffusion : comprend les activités relatives à la diffusion du livrable de veille réalisé auprès des personnes intéressées.

Afin de montrer les différentes combinaisons de fonctionnalités possibles, nous pouvons définir ce que l'on appelle le *trèfle fonctionnel des outils de veille*. Ce diagramme permet de visualiser différents types d'outils selon leur(s) couverture(s) fonctionnelle(s) (Dualé 2010).

Figure 3 : Le trèfle fonctionnel des outils de veille



(Dualé, 2010)

D'autres paramètres supplémentaires comme les fonctionnalités d'administration ou de gestion de contenus peuvent être également présents (Dualé 2010).

2.1.4.2 Modèles économiques des outils de veille

Balmisse et Meingan (2008) définissent quatre grandes familles de modèles économiques pour les logiciels de veille :

- Modèle gratuit : Accessibles librement, il s'agit le plus souvent d'outils de collecte d'informations accessibles via Internet ou monopostes.
- Modèle *open source* : Le code source du logiciel est accessible et peut être modifié et adapté.
- Modèle de licence : Il est demandé d'acheter une licence dont les tarifs peuvent varier d'un éditeur à l'autre pour utiliser le logiciel.
- Modèle abonnement : Modèle reposant sur le concept de *Software as a Service* (SaaS). L'utilisateur « loue » le service qui est hébergé hors du système d'information de l'entreprise.

2.1.4.3 Architectures des outils de veille

Du point de vue de leur architecture, trois types de logiciels de veille peuvent être distingués (Balmisse, Meingan 2008) :

- Service Internet : Comme son nom l'indique, ce type de logiciel est accessible uniquement depuis Internet.
- Logiciel monoposte : Le logiciel s'installe sur un seul ordinateur. Ainsi, seules les personnes ayant accès au poste peuvent utiliser l'outil.
- Logiciel serveur : Le logiciel est installé sur un serveur interne ou externe à l'entreprise et est accessible via un logiciel client. Il s'agit d'un outil principalement destiné à de grandes organisations.

Les différentes familles d'outils de veille

Cette segmentation en trois dimensions permet à Dualé (2010, p. 12) de mettre en évidence cinq familles d'outils « qui se distinguent par leur couverture de la problématique d'intelligence économique » :

- Les progiciels de veille : Ce type d'outils rassemblent les trois fonctionnalités détaillées précédemment, à savoir la collecte, l'analyse et la diffusion. Ils peuvent néanmoins privilégier certaines fonctions. Ils sont proposés sous contrat de licence mais également sous forme d'abonnement, selon le modèle SaaS.
- Les outils de veille métier : Ces outils sont spécialisés dans un seul type de veille bien spécifique comme la veille image ou encore la veille brevet.
- Les plateformes de gestion de contenus : Il s'agit de logiciels proposant des fonctionnalités de portail, de gestion documentaire ou encore de moteur de recherche qui peuvent être utilisés à des fins de travail collaboratif, de publication, etc. Ces outils ne sont pas exclusivement faits pour la veille mais peuvent être adaptés grâce à des paramètres spécifiques.

- Les outils spécialisés : Comme leur nom l'indique, il s'agit de logiciels dédiés à des tâches très spécifiques. Ils peuvent répondre à l'une ou plusieurs des fonctionnalités inhérentes aux outils de veille mais ne sont pas tous strictement dédiés à la veille.
- Les outils Internet : Il s'agit de service Internet souvent mis à disposition gratuitement. Ils ont pour la plupart une fonction de collecte d'information.

2.1.5 Les plateformes de veille

Une *plateforme de veille* peut être définie comme un « outil développé par des sociétés spécialisées et qui a pour but de collecter, d'analyser, synthétiser et diffuser l'information que l'on peut trouver sur Internet » (Winello 2011). Il s'agit d'une locution synonyme de l'expression *progiciel de veille* présentée précédemment par Dualé (2010).

Dans la littérature, bien que moins utilisés, on retrouve également d'autres termes allant dans le même sens comme *plateforme logicielle de veille*, *plateforme d'éditeurs de veille* ou encore *logiciel de veille intégré* (Digimind 2017).

Digimind, Sindup et MyTwip, que nous analysons dans le cadre de notre étude, font partie de cette catégorie d'outils. Tous trois fonctionnent selon le modèle abonnement et sont des services internet.

2.1.6 La notion de qualité

Selon Doucet (2010, p. 5), la qualité a deux significations principales :

1. La qualité peut être définie comme « la manière d'être, bonne ou mauvaise de quelque chose ». Il s'agit d'un état sans aucune notion de valeur. Ainsi, lorsque l'on parle de la qualité d'une plateforme de veille, celle-ci peut être bonne ou mauvaise.
2. La qualité peut également être comprise comme « la supériorité, l'excellence en quelque chose ». Dans le cas présent, une certaine notion de hiérarchie est introduite. Ainsi, un outil de veille de qualité sous-entend qu'il est meilleur qu'un autre.

Plus récemment, le concept de qualité a évolué vers une troisième définition complémentaire aux deux précédentes, introduisant les notions de produits/services et de satisfaction des clients. Dans le domaine de la veille, la qualité peut alors être définie comme étant l'aptitude de l'outil à satisfaire les besoins des utilisateurs ou à satisfaire des exigences (Cohen 2004, p. 87).

2.1.7 La notion d'évaluation

L'évaluation est un processus qui vise à déterminer la valeur d'un service, d'une installation, d'un programme, en estimant notamment son efficacité et/ou son efficience (Evaluer 2017 ; Sutter 2006). Elle peut s'appuyer sur des indicateurs mais également sur des observations ou constats (Sutter, 2006). L'évaluation repose sur des mesures qualitatives et/ou quantitatives.

2.1.8 La notion de source

Une source est définie comme étant l'origine d'une information (Sources 2017). Dans notre travail, nous utilisons la dénomination de "source" ou "source de veille" pour chacune des pages web explorées par les moteurs de recherche des plateformes de veille, qu'elles aient été insérées manuellement ou qu'elles soient automatiquement référencées dans la base de données de la plateforme.

2.1.9 La notion d'indicateurs

Selon l'Association française de normalisation (2000, p.5), un indicateur est « une information choisie, associée à un critère, destinée à en observer les évolutions à intervalles définis ». Il s'agit de la représentation quantitative ou qualitative d'un élément informationnel dans but d'en déterminer sa valeur (Cacaly, Le Coadic 2004 ; Briand et al. 2009). Un indicateur doit répondre à certains critères pour répondre à sa mission informative tels que la fidélité, l'objectivité ou encore la stabilité dans le temps, et doit toujours être lié à un objectif précis (Cacaly, Le Coadic 2004).

2.1.10 La notion de résultats

Dans le cadre de notre étude, les résultats représentent les URL renvoyées par les plateformes de veille, en réponse aux requêtes insérées. Ces résultats peuvent être de trois catégories : pertinents, non pertinents ou non analysables (cf. chapitre 3.5.3 *Choix des indicateurs*).

Ils sont donc à distinguer des résultats de la veille en tant que produits ou livrables de veille, qui représentent des "documents émis périodiquement ou à la demande qui font la synthèse des recherches ponctuelles ou régulières sur des thèmes ou sur des concurrents selon les besoins d'un groupe d'utilisateurs" (Bachr 2008).

2.2 Les outils de veille

2.2.1 Les plateformes de veille évaluées

2.2.1.1 Digimind

La société et ses offres

Fondée en 1998 par Paul Vivant à Grenoble, la société Digimind est implantée à Paris, Singapour, en Amérique du Nord et en Afrique, et dénombre une centaine d'employés. Par ailleurs, son chiffre d'affaire, en croissance continue depuis sa création, s'élevait à 10 millions d'euros en 2014 (Schwob 2017a).

L'offre de cette société, variée, permet de couvrir l'ensemble des étapes de la veille, depuis la collecte jusqu'à la diffusion de l'information.

La première solution développée par la société, Digimind Intelligence, se destine aux cellules de veille et spécialistes de l'information et permet la surveillance de tout type de sources. Afin de s'adapter aux dernières innovations du marché, Digimind a modernisé cette solution en intégrant diverses pratiques et fonctions sociales issues des réseaux sociaux, et ce afin de proposer une interface utilisateur plus conviviale, plus propice à la collaboration, en somme plus en adéquation avec les habitudes et les besoins des utilisateurs (Asselin 2015).

Par ailleurs, Digimind a lancé Digimind Social Analytics en 2013. Développée pour les personnes du marketing et de la communication, la plateforme permet aux entreprises de gérer l'ensemble de leurs comptes sur les réseaux sociaux, mais également d'analyser la communication de leurs concurrents (Ferrière 2015). A cela s'ajoute la possibilité d'analyser l'e-réputation, de mesurer le retour sur investissement des campagnes marketing ainsi que de créer des rapports (Schwob 2017a).

Typologie des sources surveillées

L'outil permet de surveiller tout type de sources, parmi lesquelles nous pouvons citer :

- L'ensemble des sources que l'on peut trouver sur le web,
- Les réseaux sociaux (Twitter, Facebook, Instagram, Pinterest, LinkedIn, Vine),
- Les forums,
- Les blogs,
- Les bases de données,
- Les sites d'actualités,
- Les vidéos (Editeurs de solutions de veille et e-réputation 2015).

A cela s'ajoute également la reconnaissance d'images depuis 2016, ce qui représente un point important pour assurer une veille en réputation, notamment concernant la reconnaissance des logos (Sala 2016).

Points forts et points faibles

La société Digimind est reconnue comme étant le leader européen sur le marché de la veille. Sa place a été confirmée par l'étude 2016 de G2Crowd (interface qui recueille les avis des utilisateurs de logiciels et qui analyse et mesure la performance des outils de monitoring). Parmi les 34 outils de veille comparés, Digimind est la solution qui obtient le taux de satisfaction des utilisateurs le plus haut, tout en occupant la place la plus importante sur le marché, malgré une très forte concurrence (Rebmann 2016).

Ainsi, s'il est difficile de soulever des points faibles à propos de cette plateforme très complète, il est tout de même possible de critiquer un abonnement élevé (plusieurs dizaines de milliers d'euros), non accessible aux PME, contrairement à d'autres solutions. C'est pourquoi la majeure partie de ses clients sont de grands groupes telles que Deloitte, Nestlé, SFR, Siemens, Schneider Electric, Toyota, Intel, Alstom, etc.

2.2.1.2 Sindup

La société et ses offres

Plus récente que Digimind, la société française Sindup est créée en 2005 par Mickaël Réault, analyste développeur de formation, et commercialise son premier outil de veille en 2009. Tout comme Digimind, cette plateforme surveille non seulement les sources classiques, mais également les médias sociaux.

Depuis 2012, suite à une stratégie portée vers l'international, la plateforme est multilingue et le sourcing mondial (Sindup [n.c. a]). Présente sur toute la francophonie, son offre touche aussi le marché d'Amérique du Nord et du Brésil.

Afin de s'adapter aux besoins et aux budgets de ses différents clients, la solution Sindup se décline en trois offres (Sindup [n.c. b]). Par ailleurs, deux produits sont proposés en Saas par cette société : Sindup et Detectup (Schwob [n.c.]). Si Sindup est un outil comparable à la solution Digimind Intelligence, Detectup est quant à lui un service clé en main, externalisé, à destination des PME et TPE qui n'ont pas les moyens humains et/ou financiers de développer une cellule de veille à l'interne (Sindup, l'étoile montante de la veille [n.c.]).

Typologie des sources surveillées

L'outil permet de surveiller différents types de sources :

- Les actualités (sites de presse, institutionnels, d'entreprises, etc.),
- Les avis de consommateurs (type TripAdvisor),
- Les médias sociaux (les pages Facebook, Twitter, Google+, Pinterest, Instagram, YouTube, Dailymotion ainsi que les blogs),
- Les forums,
- L'Inbox (mails, pièces jointes, informations informelles...),
- Les documents (Word, Excel, PDF, PPT).

Par ailleurs, la plateforme surveille non seulement les sources gratuites, mais peut également surveiller les sources payantes si le client possède un abonnement (Tisserand-Barthole 2015).

Ses points forts et ses points faibles

Contrairement à Digimind, Sindup possède l'avantage de proposer des services à prix abordable pour les entreprises et autre institution possédant un budget limité. Par ailleurs, ses logiciels, facilement paramétrables, possèdent une interface simple et attractive, et les tableaux de bord proposent une analyse automatique des résultats. Enfin, en adéquation avec l'évolution du marché ces dernières années, les produits permettent d'effectuer une veille sur les réseaux sociaux.

Cependant, il apparaît que le système est particulièrement lent, les fonctionnalités restreintes en comparaison de plateformes plus importantes, et ses algorithmes de filtrage perfectibles (car laissant passer beaucoup de bruits) (Schwob [n.c.]).

2.2.1.3 MyTwip

La société et ses offres

MyTwip, qui se définit comme un “moteur professionnel de veille thématique et sectorielle”, est une solution développée en 2007 par la société française Coexel. Son président, Vincent Boisard, ingénieur de formation, a orienté l’outil vers les secteurs technologiques innovants (Coexel 2017a). Par ailleurs, cette société s’est spécialisée en analyse sémantique et en big data (Schwob 2017b).

Tout comme Sindup, afin de s’adapter aux besoins de ses différents clients, elle a développé une solution modulable selon trois profils types (Coexel 2017a).

En complément de sa solution, Coexel accompagne ses clients dans la définition de leurs besoins en information et dans l’appropriation des données collectées.

Typologie des sources surveillées

Cette solution permet de surveiller toute source accessible en ligne, avec ou sans mot de passe (Editeurs de solutions de veille et e-réputation 2015).

Il s’agit, entre autres :

- des sources issues du web,
- des flux RSS,
- des blogs,
- des publications scientifiques, brevets et appels d’offres,
- de Twitter,
- des newsletters et e-mails,
- des formulaires et bases en ligne (avec login et mot de passe) (Boisard 2016).

Ses points forts et ses points faibles

Grande base de sources, elle a la particularité de permettre la veille sur les brevets. Si ses algorithmes de recherche sont très puissants, le temps d’importation des sources est cependant relativement long (entre 24h et 72h) (Schwob 2017b).

Par ailleurs, contrairement à d’autres plateformes, de nombreux paramétrages ne sont pas réalisables côté utilisateurs, et nécessitent d’être effectués par la société elle-même.

2.2.2 Les comparatifs d’outils de veille

Comme nous avons pu l’expliquer précédemment, afin d’analyser les résultats retournés par les trois plateformes de veille étudiées, nous devons au préalable définir les indicateurs permettant la réalisation de cette évaluation.

Pour cela, il nous a semblé nécessaire de procéder à un état de l'art des comparatifs déjà existants à ce sujet, afin de relever les critères utilisés habituellement pour procéder au choix d'un outil de veille.

Archimag, magazine spécialisé à destination des professionnels en information documentaire, publie régulièrement un panorama des principaux logiciels de veille disponibles sur le marché. Ces benchmarks permettent de comparer les fonctionnalités proposées par les différentes solutions, ainsi que les prix pratiqués.

Le dernier panorama, datant de juin 2017 (Panorama de solutions de veille 2017) et orienté veille collaborative, permet de comparer 10 logiciels de veille, dont MyTwip et Digimind Intelligence. La comparaison s'effectue au niveau des options proposées par les différents logiciels :

- la catégorisation automatique des données collectées,
- l'analyse de la tonalité,
- la création de résumés automatiques,
- les différentes langues supportées,
- les différents types de visualisations des données proposés,
- les différents modes de diffusion de la veille possible,
- les outils de statistiques mis à disposition.

Les critères de comparaison de ce panorama sont très similaires au benchmark diffusé fin 2015 (Editeurs de solutions de veille et e-réputation 2015) à propos des solutions de veille et e-réputation. Quelques spécifications techniques supplémentaires comparées étaient cependant à relever :

- le système d'exploitation supporté (Windows, Linux et Mac, mais aussi les différents navigateurs web),
- le type d'architecture (Saas, serveur, monoposte, etc.),
- les types de sources pouvant être surveillées (sites web, blogs, réseaux sociaux, sources payantes, etc., mais aussi les formats supportés, à savoir : texte, image, vidéo...),
- la présence ou pas de la technologie "data mining", et ce que cela représente.

Enfin, il est possible de relever un dernier panorama des outils de veille (Solutions dans la course à l'innovation 2012), issu du guide pratique 47 traitant des outils et de l'efficacité des systèmes de veille. Ce benchmark utilise des critères beaucoup plus précis que ceux de 2015 et de 2017, toujours en lien avec les fonctionnalités des outils. Nous pouvons entre autres citer :

- la vérification des modifications sur les pages des sites web,
- la fonction de classement par clustering,
- l'extraction de paragraphes pertinents,

- l'analyse du discours par le biais de la tonalité (e-reputation) (Solutions dans la course à l'innovation 2012).

Il serait inutile de remonter plus loin dans les comparatifs publiés régulièrement par ce magazine, car les plateformes de veille évoluent extrêmement rapidement, et des tableaux plus anciens seraient peu profitables à notre recherche.

Nous pouvons relever, en conclusion, que les auteurs de ces benchmarks s'intéressent plus aux fonctionnalités proposées par les différentes solutions qu'à la qualité des informations retournées à proprement parler, bien que certains critères soient essentiels pour mesurer la performance d'une plateforme (les sources surveillées, les formats et langues supportés, le data mining, etc.).

Un autre comparatif de logiciels, paru en début d'année 2017, a été particulièrement diffusé et partagé dans le monde de la veille. Il s'agit d'une infographie dynamique créée par le club Webtechno AEGE (Guillaume Sylvestre et Antoine Violet-Surcouf) à partir de données publiques disponibles.

L'outil permet de comparer en ligne 44 outils de veille gratuits et payants selon plusieurs paramètres :

- la date de création et le nom de l'outil,
- le(s) type(s) d'alerte proposé(s) (e-mail, RSS, etc.),
- l'installation (Saas, desktop ou extension d'un navigateur),
- le coût (gratuit, payant, freemium), et s'il est payant, à combien s'élève la dépense,
- la prise en main administrateur (rapide, modérée ou technique),
- le type de veille (réseaux sociaux, agent d'alerte, plateforme de veille, etc.),
- la spécialité de l'outil (réseaux sociaux, brevets, news, etc.) (Club Webtechno AEGE 2017).

Très ludique, cette infographie permet de répondre aux besoins des utilisateurs en délimitant les outils selon leur tarif et leurs caractéristiques principales.

Nous pouvons, pour continuer ce panorama non exhaustif des comparatifs de plateformes de veille, également citer le guide réalisé par la société Digimind (Digimind 2011), qui a pour objectif d'aider les chefs de projet chargés de choisir et de mettre en place un tel outil.

Bien que ce document date de 2011 et soit écrit par l'un des acteurs du marché (ce qui représente un biais non négligeable - l'outil de Digimind répondant positivement à tous les points présentés), il apparaît comme relativement complet. En effet, deux cents critères sont présentés et permettent de faire un choix au regard de l'ensemble des fonctionnalités possibles, couvrant les 6 étapes du processus de veille (Digimind 2011).

S'il n'est pas possible ici de tous les énumérer, quelques-uns méritent cependant d'être relevés :

- l'intégration du score d'influence des sources : ce critère correspond à une évaluation automatique et manuelle de l'influence de la source sur un sujet donné.

- la précision de la collecte : sans identifier de critère particulier, la plateforme de veille doit permettre une recherche bien plus efficace que les moteurs généralistes tels que Google. Pour cela, il est expliqué que les technologies proposées par l'outil doivent permettre de supprimer au mieux le bruit informationnel sans générer non plus le silence.
- l'exhaustivité de la surveillance : les outils de veille doivent permettre une surveillance exhaustive, en récupérant toutes les informations existantes quel que soit le format, l'encodage, etc. Ainsi, plus une plateforme proposera de fonctionnalités, plus elle sera capable d'être exhaustive concernant les sources qui nous intéressent.
- l'analyse et le text-mining : les outils de veille doivent permettre d'homogénéiser et de structurer le grand nombre d'informations collectées. Cela nécessite des fonctions telles que la reconnaissance et l'extraction automatique d'informations (Digimind 2011).

Avec le même objectif que le guide rédigé par la société Digimind, nous pouvons aussi relever le document réalisé par les Editions Weka à propos des critères de choix d'un outil de veille (Dualé 2010). Dans le cadre d'un projet de veille, le choix de la plateforme est crucial, et ce document, en recensant l'ensemble des fonctionnalités potentielles, doit permettre de choisir l'outil qui soit le plus à même de répondre aux besoins de l'organisation.

Bien sûr, on y retrouve la description de nombreuses options qui apparaissent comme essentielles telles que :

- la gestion des sources d'information externes (obtenues en dehors des sources électroniques),
- l'ajout de commentaires,
- le classement manuel de l'information dans des catégories,
- la recherche sur plusieurs types de sources d'informations,
- le choix des langues supportées,
- les types de représentations visuelles de l'information utilisées,
- la présence d'outils d'aide à la création de la newsletter,
- une interface utilisateur personnalisable,
- la gestion des profils et des droits (avec la possibilité de création de nouveaux rôles, la granularité de la gestion des droits d'accès),
- etc.

Un critère retient en particulier notre attention. Nommé "montée en charge", ce dernier traite du temps de réponse de la plateforme, et permet de juger la performance de la collecte et de l'indexation.

Enfin, pour terminer cette vue d'ensemble des comparatifs, deux articles ont retenu notre attention. Il s'agit de retours d'études dont l'objectif était de comparer deux à trois logiciels de

surveillance du web : Talk Walker Alerts, Google Alerts et Mention (Khakimov 2013 ; Woodward 2013).

Ces études sont intéressantes, car proches de la nôtre pour ce qui est de la méthodologie. En effet, sur la base de mêmes requêtes et sur un temps donné, les résultats des différents logiciels sont prélevés. De cette manière, les outils peuvent ensuite être comparés selon certains critères, qui reposent sur trois grands axes :

- la quantité d'alertes délivrées,
- la qualité des alertes,
- et la rapidité d'envoi (par rapport à la publication de l'information sur le net).

Concernant la quantité d'alertes délivrés, les auteurs prennent en considération :

- le nombre d'emails reçus (ou alertes - qui sont des notifications de nouveaux résultats obtenus),
- et le nombre de liens collectés (présents dans les emails).

A propos de la "qualité des alertes", ils ont examiné :

- le nombre de résultats identiques (ou au contraire uniques) aux différents outils,
- la duplication des alertes envoyées par un même logiciel (même URL envoyée à plusieurs reprises),
- le taux de présence des mots-clés dans les résultats obtenus (dans le titre, la meta description ou dans la page elle-même),
- et la qualité des sites dont les informations sont issues (l'auteur cite la fiabilité du site ou encore la variété des sources - Khakimov 2013).

Un des auteurs considère également le nombre d'URL indexées par Google comme critère de qualité (Woodward 2013).

2.3 L'évaluation de la recherche d'information

Comme nous l'avons vu dans notre cadre conceptuel, la recherche d'information (RI) est une notion directement liée à la veille. Ainsi, l'évaluation de la qualité d'une plateforme de veille peut s'inspirer de la discipline de l'*Évaluation de la recherche d'information*, appelée également, *Évaluation des systèmes de recherche d'information*.

2.3.1 Recherche d'information et système de recherche d'information

Un système de recherche d'information (SRI) est un outil, généralement un moteur de recherche, qui a pour objectif de retrouver les documents pertinents à une requête d'un utilisateur, à partir d'une collection de documents (Nie 2016a). Cette définition fait appel à différentes notions clés :

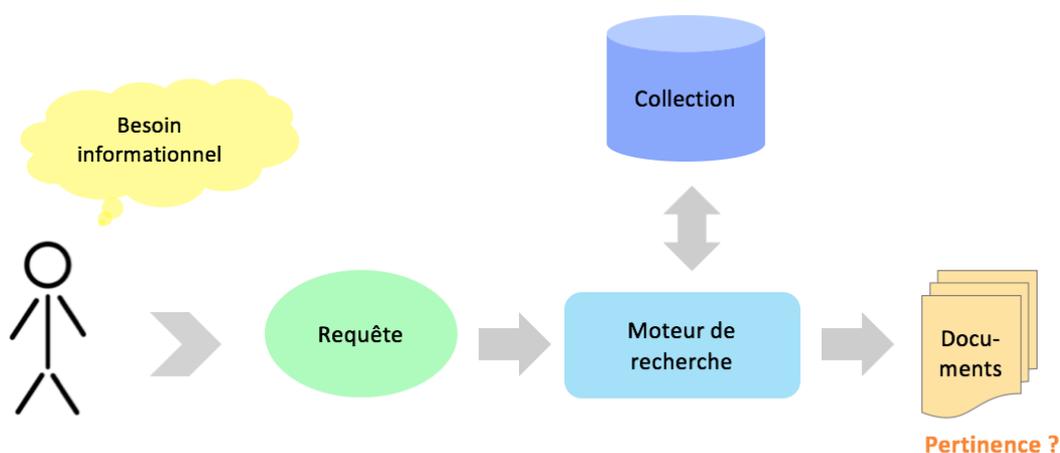
Le document : il s'agit de tout élément susceptible de répondre à la requête d'un utilisateur (texte, vidéo, image, page web, ...) (Nie 2016a).

La requête : elle se compose d'un ou plusieurs mots-clés combinés et vise à exprimer le besoin informationnel d'un utilisateur (Desfriches Doria 2015). Dans la réalité, elle n'est bien souvent qu'une description partielle de son besoin d'information. En effet, il est très difficile, voire même impossible, de formuler une requête qui le décrit de manière précise et complète (Nie 2016a).

La collection : il s'agit d'une base de documents telle que le web ou le catalogue d'une bibliothèque par exemple (Gobeill 2016).

La pertinence : il s'agit de la relation entre le besoin en information et les documents attendus. Un document pertinent est un document dans lequel l'utilisateur peut trouver l'information dont il a besoin (Nie 2016a). A contrario, un document est jugé non pertinent s'il ne répond pas aux attentes de l'utilisateur. La pertinence est une notion très complexe et variable puisqu'elle repose uniquement sur la subjectivité de la personne qui l'évalue (Audeh et al. 2015). Ainsi, pour une même requête, un document retourné peut être considéré comme pertinent par un utilisateur et non pertinent par un autre, l'évaluation ne se faisant pas sur les mêmes critères (Gobeill 2016).

Figure 4 : Le modèle RI



(Inspiré de Gobeill, 2016)

Nous pouvons considérer deux grandes catégories de SRI. Les SRI faisant du ranking, c'est à dire que les documents sont classés et référencés selon des critères et algorithmes qui leurs sont propres, pour proposer les résultats les plus pertinents aux premiers rangs. C'est notamment le cas des moteurs de recherche comme Google, qui définissent le rang de chacun des sites web qu'ils recensent. A contrario, certains SRI ne font pas de ranking. Ainsi, les résultats ne sont pas ordonnés par pertinence mais sont présentés de manière aléatoire ou en suivant certains critères. C'est notamment le cas des trois plateformes de veille que nous évaluons, qui ordonnent les résultats selon leur date de réception.

2.3.2 Évaluation d'un SRI

L'Évaluation de la RI s'intéresse à évaluer l'efficacité des SRI, c'est-à-dire leur capacité à retourner des documents pertinents (Pinel-Sauvagnat, Mothe 2013). Le modèle d'évaluation de référence dans le domaine est le projet *Cranfield* dirigé par Cyril Cleverdon dans les années 1960 (Nie 2016b). Bien que d'autres projets dans le domaine aient vu le jour depuis, *Cranfield* reste l'approche dominante sur laquelle l'évaluation de SRI s'appuie (Baccini et al. 2010). Le scénario actuel le plus connu est celui de la recherche *ad hoc* (Pinel-Sauvagnat, Mothe 2013).

Ce modèle d'évaluation repose sur des collections de test, aussi nommées collections de référence (Pinel-Sauvagnat, Mothe 2013). Ces corpus sont composés de :

- Une collection de documents : Il faut avoir un nombre de documents assez élevé pour qu'un corpus de texte soit significatif. Les collections de test récentes peuvent comprendre 100'000 comme des millions de documents (Nie 2016a).
- Un ensemble de requêtes : Les requêtes sont composées d'un ou plusieurs mots-clés et visent à traduire les besoins informationnels des utilisateurs. Idéalement, il faut réaliser quelques dizaines de requêtes traitant de sujets variés (Nie 2016a)
- La liste de documents pertinents pour chaque requête : Il s'agit, pour chaque requête, d'une liste des documents répondant au(x) besoin(s) informationnel(s) de l'utilisateur. Pour établir cette liste, la pertinence de chaque document de la collection doit être évaluée selon chacune des requêtes (Nie 2016a).

Afin d'évaluer la qualité d'un SRI, on procède de la façon suivante : le moteur de recherche exécute les requêtes les unes après les autres sur la collection de documents et renvoie pour chacune une liste ordonnée de documents qu'il considère comme pertinents. Différentes mesures d'efficacité sont ensuite appliquées et calculées (Pinel-Sauvagnat, Mothe 2013)

2.3.3 Les métriques de la RI

Il existe de nombreuses métriques permettant d'évaluer un SRI de manière précise. Le programme d'évaluation *trec_eval*, notamment utilisé dans le programme *TREC* (Text REtrieval Conference), propose 135 mesures (Baccini et al. 2010). Un grand nombre d'entre elles découlent de deux notions historiques issues du modèle *Cranfield*, la précision et le rappel (Audeh et al. 2015).

En comparaison aux SRI faisant du *ranking*, peu de métriques peuvent être utilisées pour évaluer des SRI ne faisant pas de *ranking*. Les plus importantes sont les deux citées précédemment : la précision et le rappel (Manning et al. 2008).

La précision et le rappel

Une requête faite à un SRI permet de diviser la base de documents en 4 ensembles (Lafouge et al. 2002 ; Gobeill 2016) :

- TP (True positives) : les documents retournés pertinents
- FP (False positives) : les documents retournés non pertinents
- FN (False negatives) : les documents pas retournés pertinents
- TN (True negatives) : les documents pas retournés non pertinents

Tableau 1 : Tableau de contingence des documents

| Résultats | Retournés | Non retournés |
|----------------|-----------|---------------|
| Pertinents | TP | FN |
| Non pertinents | FP | TN |

(Inspiré de Lafouge et al. 2002 ; Gobeill 2016)

Précision (P) : Mesure de l'efficacité d'un SRI en comparant le nombre de documents pertinents retournés par rapport au nombre total de documents retournés. Dans un langage propre aux Sciences de l'information, cette métrique mesure le bruit documentaire (Desfriches Doria 2015).

$$P = \frac{TP}{TP + FP}$$

Rappel (R) : Mesure de l'efficacité d'un SRI en calculant le ratio entre le nombre de documents pertinents retournés et le nombre total de documents pertinents existants dans la collection du système. Il s'agit d'un indicateur permettant de mesurer le silence documentaire (Desfriches Doria 2015).

$$R = \frac{TP}{TP + FN}$$

Ces deux métriques sont très fortement corrélées. Ainsi, lorsque l'une d'entre elle augmente, l'autre a tendance à diminuer. En effet, pour avoir le meilleur taux de rappel (qui est de 1), un SRI pourrait retourner tous les documents contenus dans la collection. Néanmoins, dans ce cas-ci, la précision serait très basse. A contrario, pour atteindre le degré maximal de précision, le SRI pourrait retourner très peu de documents en réponse. Mais dans le cas présent, le rappel en souffrirait (Nie 2016a).

3. Méthodologie de la recherche

Dans ce chapitre, nous nous attachons à décrire les méthodes que nous avons choisies afin de mener à bien notre étude. Dans un premier temps, nous expliquerons l'approche générale adoptée. Nous détaillerons ensuite la méthodologie suivie pour mettre en place notre processus de veille et ainsi collecter nos données. Finalement, nous décrirons les critères choisis pour comparer et évaluer les résultats des trois plateformes de veille et la méthode appliquée pour les analyser.

3.1 Approche générale

Comme nous l'avons vu dans notre revue de littérature, les plateformes de veille fonctionnent de la même façon que les moteurs de recherche. L'utilisateur entre une requête et des résultats, contenus dans une base de données, lui sont retournés. En plus d'interroger leur propre base de données, les plateformes de veille permettent également à l'utilisateur d'intégrer des sources qu'il a lui-même choisies.

Dès lors, nous avons décidé de mener notre étude en prenant en compte ces deux volets. Dans un premier temps, nous avons souhaité étudier les résultats retournés lorsque les plateformes interrogent leurs propres sources (que l'on appellera "sources automatiques" ou SA) et d'autre part, comparer les résultats retournés par les trois plateformes lorsque celles-ci interrogent des sources identiques, sélectionnées et ajoutées manuellement (que l'on nommera "sources manuelles" ou SM).

Dans les deux cas, pour permettre la comparaison, les plateformes ont été paramétrées avec les mêmes requêtes, bien que la syntaxe ait été adaptée à chacune d'entre elles. Des indicateurs ont ensuite été définis de façon à comparer les résultats et à en tirer des conclusions.

3.2 Le ciblage

3.2.1 Analyse des besoins et recherches exploratoires

Pour mener à bien notre étude et générer des données à analyser, nous avons mis en place une veille de quelques semaines suivant les étapes habituelles du processus de veille (cf. chapitre 2.1.2 *Le cycle de la veille*).

Pour rappel, le choix des trois sujets de veille a été fait par la Directrice de ce travail, Hélène Madinier, selon ses propres besoins et ceux de mandants externes. Les trois sujets de veille de notre projet sont les suivants (cf. Annexe 1) :

- 1) Dans le domaine des semi-conducteurs de puissance : une veille concurrentielle (pour laquelle les différents concurrents ont été définis en amont), un suivi des investissements et des fusions-acquisitions des acteurs-clé du marché (également selon une liste préalablement définie) ainsi que l'évolution du marché des semi-conducteurs de puissance aux USA.
- 2) Dans le domaine des Sciences de l'information : l'évolution des métiers en Suisse et dans les pays voisins, avec focalisation sur les offres d'emploi, l'identification des attentes et le suivi de l'évolution des besoins.

3) Dans le domaine de l'loT (ou *industrie 4.0*) : obtenir des informations sur cette technologie adaptée au domaine de la santé et de l'industrie pharmaceutique (en veillant sur les fournisseurs, les clients, les pratiques), essentiellement en Suisse.

La mise en place d'une veille nécessite en amont la réalisation de quelques bonnes pratiques. Comme nous l'avons expliqué, la première phase du cycle de veille est l'étape du ciblage. Celle-ci consiste notamment en la définition du périmètre de veille et en la formalisation des besoins (Université européenne de Bretagne 2016). En contact direct avec les différents mandants, notre Directrice, Hélène Madinier, s'est chargée d'une partie de cette phase du processus et nous a transmis les diverses informations dont nous avons besoin.

Pour des questions de faisabilité et d'un commun accord, le sujet de veille concernant les Sciences de l'information a été restreint géographiquement à la Suisse et à la France, qui sont les pays les plus importants aux yeux du mandant. Par ailleurs, pour ces mêmes raisons, mais de manière à garantir tout de même une certaine exhaustivité, il a été décidé de limiter la rédaction des requêtes au français et à l'anglais.

Afin de mieux délimiter et comprendre leurs enjeux, nous avons également mené des recherches exploratoires sur les trois sujets. Nous avons principalement utilisé le moteur de recherche Google en y insérant diverses requêtes en français et parfois, en anglais. Cette étape a aussi été l'occasion de relever les sources pertinentes traitant de ces sujets (Schwob 2017).

3.2.2 Plans de veille

Le plan de veille est un tableau qui synthétise les besoins définis précédemment. Il les formalise en thèmes et sous-thèmes (axes et sous-axes), décrit les objectifs qui y sont liés et recense les types de sources pertinentes à collecter. Il peut également donner des éléments de délai ou de responsabilités (Madinier 2016).

Nous avons réalisé, pour chacun de nos sujets de veille, un plan de veille sous forme de tableau (cf. Annexe 2), dont les trois colonnes permettent de préciser :

- les différents axes pour chacun des sujets de veille,
- les enjeux, associés à chacun de ces axes,
- les types de sources permettant de collecter les informations souhaitées (l'étape de l'évaluation des besoins ayant permis dans un premier temps de relever les sources susceptibles d'être pertinentes).

Déterminer les enjeux associés à chacun des axes de veille est essentiel et correspond à l'«identification des fonctions attendues». En effet, il faut être en mesure d'expliquer pourquoi il est intéressant de suivre un sujet, sous quel angle, quels sont les avantages recherchés ainsi que les profits attendus (Frochot 2006).

Cette étape a nécessité en amont de mieux cerner les différents types de veille existants ainsi que leurs objectifs :

- la veille concurrentielle va permettre de mieux connaître les mouvements des concurrents (les recrutements, les acquisitions, les nouvelles marques, etc.) et d'évaluer leur positionnement par rapport à l'entreprise mandante, en surveillant notamment leurs actualités dans la presse locale, nationale et spécialisée. Il s'agit non seulement de s'informer sur leurs prestations, leurs produits, mais aussi sur leurs clients et sur leurs prix ;
- la veille métier (ou veille sectorielle) va quant à elle permettre de connaître et d'anticiper les évolutions du marché, identifier les menaces et opportunités, les leviers de croissance, les activités en perte de vitesse. Cela requiert entre autres de surveiller l'actualité dans la presse locale, nationale et spécialisée, mais aussi de s'informer sur les tendances comportementales et sociétales ;
- enfin, la veille technologique a pour objectif d'appréhender les innovations et évolutions du marché, de manière à anticiper les mutations technologiques du marché et connaître les mouvements des concurrents. Pour cela, il est nécessaire de surveiller la presse et les sites d'informations spécialisés, les dépôts de marque, dessins ou modèles, brevets de ce marché, ainsi que les nouvelles publications et travaux de recherche (L'Œil au Carré 2013).

3.2.3 Pré-test

Durant la phase de ciblage, nous avons également procédé à un pré-test sur un sujet d'actualité afin d'évaluer si notre approche générale était pertinente. En effet, en insérant exactement les mêmes sources dans les trois plateformes, nous appréhendions que ces dernières nous renvoient exactement les mêmes données, et que de ce fait, notre démarche concernant la comparaison des résultats des "sources manuelles" soit inutile. Néanmoins, les plateformes n'ont pas collecté les mêmes résultats, ce qui a prouvé la pertinence de notre approche. Ce test nous a également permis d'apprendre à manipuler les différentes plateformes.

3.3 Sourcing et plans de veille

3.3.1 Le choix des sources

Le sourcing, partie intégrante du cycle de veille, est une approche qui consiste à établir une base de sources pertinentes via des recherches ponctuelles (Delengaigne 2012).

Une des possibilités pour identifier les sources nécessaires est de simplement lancer une recherche sur Google à l'aide de quelques mots-clés. Pour compléter cette démarche, et de manière plus efficace, il est également essentiel de rechercher des pages de liens (rédigées par des experts ou des associations) permettant d'identifier rapidement les sources pertinentes. Pour cela, des blogs, des sites de bookmarking social, de microblogging ou des plateformes de curation peuvent se révéler utiles (Foenix-Riou 2014).

Notre étude ne visant pas une veille exhaustive sur les trois sujets mais une analyse comparative de trois plateformes de veille, nous avons sélectionné des sources qui peuvent être surveillées par de tels logiciels. Ainsi, nous n'avons, par exemple, pas paramétré d'alertes e-mail, ce que nous aurions fait dans le cas d'un réel processus de veille.

Le choix des sources a donc été établi de manière à répondre à nos besoins, mais certaines limites ont dû être posées pour des questions de faisabilité. Dans le cadre de la recherche sur l'évolution du marché de l'emploi dans le domaine des Sciences de l'information, il n'était par exemple pas possible d'établir une veille sur l'ensemble des sites web informant des nouvelles offres d'emploi, ou sur l'ensemble des sites institutionnels de toutes les communes suisses et françaises. Un choix (vers les sites les plus connus ou les communes les plus importantes) a donc été fait. De plus, certaines sources proposent uniquement un moteur de recherche et ne peuvent donc pas être surveillées par une plateforme de veille. Toujours dans un souci de faisabilité, de manière à limiter le nombre de sources à suivre, lorsqu'un flux RSS existait, il a été préféré au suivi de la page.

Par ailleurs, certaines limites ont été posées par les plateformes, nos versions ne possédant pas l'ensemble des fonctionnalités possibles. Ainsi, l'outil de Coexel, MyTwip, ne nous a, par exemple, pas permis de mener nos différentes veilles sur les réseaux sociaux, bien que ceux-ci auraient pu nous permettre de collecter des résultats pertinents.

Enfin, afin d'éviter le risque de silence documentaire, nous avons pris soin de sélectionner non seulement des sources fréquemment actualisées, mais aussi au sein de ces sources la page la plus souvent mise à jour (les plateformes n'étant pas en mesure de surveiller l'ensemble des pages d'un site web).

3.3.2 Plans de recherche

Si le plan de veille permet de définir les besoins en information et précise les orientations stratégiques nécessaires, le plan de recherche permet quant à lui de rechercher et de trouver ces informations (Madinier 2016).

Le plan de recherche (aussi appelé plan de veille détaillé), va venir préciser le plan de veille, en décrivant :

- la problématique (à savoir les axes ou thèmes),
- la liste des sources qui seront surveillées (le sourcing),
- les moyens pour surveiller ces sources, que ce soit la/les personne-s qui s'en chargent comme les outils à utiliser,
- les termes (ou mots-clés) ainsi que les requêtes (Madinier 2016).

Nos plans, en trois colonnes, décrivent :

- les axes détaillés dans les plans de veille,
- la liste des sources à surveiller, regroupées par types,
- et enfin des notes quant aux choix des sources à suivre (précisant certains choix ou certaines restrictions pour des questions de faisabilité) (cf. Annexe 3).

Nous n'avons pas estimé nécessaire de préciser par une colonne supplémentaire les moyens nécessaires pour la surveillance de ces sources, puisque, concernant les outils utilisés, chaque source doit être surveillée par les trois plateformes de veille.

Cependant, une répartition du travail a tout de même été effectuée afin de faciliter le suivi de la collecte :

- le sujet sur l'évolution des métiers en Sciences de l'information, nécessitant un développement conséquent, a été assuré par un seul veilleur,
- les sujets sur Lemsys (semi-conducteur de puissance) et sur l'IoT appliqué au domaine de la santé ont été suivis par le deuxième veilleur.

A propos des mots-clés et requêtes, pour une meilleure lisibilité, nous avons créé des documents à part des plans de recherche, recensant l'ensemble des requêtes établies (cf. Annexe 5). Comme le préconise Didier Frochot (2006), nous avons sélectionné nos mots-clés après les avoir testés sur des moteurs de recherche, en les combinant de manière à établir des pré-requêtes, cela afin de garantir leurs bonnes formulations et la récolte des informations attendues lors de la période de veille.

De plus, afin de permettre une meilleure visualisation des différents mots-clés sélectionnés et de faciliter le choix de termes supplémentaires (synonymes, mots proches, etc.), nous avons également établi des cartes mentales (ou mind mapping en anglais), tel que le suggère Xavier Delengaigne (2012, p. 32) (cf. Annexe 4).

3.4 Paramétrage des plateformes

Le paramétrage des plateformes de veille a pu se réaliser en parallèle de la rédaction des plans de recherche. En effet, nous ne pouvions nous contenter de tester les diverses requêtes dans des moteurs de recherche généralistes. Il était au contraire important que nous puissions les tester directement dans les outils afin d'analyser leurs résultats et de les affiner si nécessaire. C'est ainsi que certaines requêtes rapportant trop de bruit ont dû être précisées par l'ajout de mots-clés ou de conditions, tandis que d'autres, nous rapportant du silence, ont nécessité d'être simplifiées.

Cette phase de test a également permis de constater parfois la nécessité de modifier le langage utilisé dans nos requêtes de manière qu'elles soient comprises par les plateformes. Il s'est par exemple révélé que la plateforme Digimind ne comprenait pas les termes mis entre guillemets, si ces derniers n'étaient pas insérés manuellement (et non copiés depuis un autre document).

A cela s'ajoute que les trois plateformes n'utilisent pas nécessairement les mêmes règles pour la rédaction des requêtes complexes. Si Digimind et Sindup utilisent les booléens en anglais, il a fallu les traduire en français pour la plateforme MyTwip. De plus, cette dernière ne comprend pas l'usage de l'astérisque pour les troncatures (aucun symbole n'est au contraire nécessaire). Par ailleurs, une requête complexe ne peut excéder vingt mots-clés dans Sindup. De façon à permettre la comparaison avec les autres outils, nous avons fait en sorte qu'aucune requête n'excède cette limite. Cela a parfois nécessité de rédiger la requête en plusieurs parties (en fragmentant les listes de concurrents pour le sujet Lemsys (semi-conducteur de puissance), ou en rédigeant la requête avec des mots-clés en français, puis dans un second temps avec des mots-clés en anglais) (cf. Annexe 5).

Il est à noter que MyTwip est une plateforme qui se paramètre de façon particulière en comparaison de Digimind et de Sindup. En effet, au préalable de la rédaction des requêtes à appliquer sur la base de données de la plateforme (ce que nous nommons par commodité “sources automatiques”), il est nécessaire de paramétrer la thématique à disposition (qui s’apparente à un fonds documentaire) dans laquelle seront déversés les résultats. La thématique peut recouvrir plusieurs sujets, dont les divers résultats peuvent ensuite être classés dans différents dossiers.

Afin de paramétrer la thématique, il faut lui appliquer un filtre (en rapport avec le sujet de la veille) qui soit suffisamment générique de manière à éviter tout silence (les résultats étant ensuite précisés par les diverses requêtes). Pour notre projet de recherche, nous avons demandé que soit créée une thématique par sujet de veille.

Le paramétrage de la plateforme MyTwip est également singulier quant aux requêtes à appliquer sur les sources entrées manuellement (identifiées et sélectionnées par le veilleur). S’il est possible de sélectionner une source et de lui appliquer une requête particulière, il est en revanche impossible de sélectionner un bouquet de sources afin de lui appliquer une ou plusieurs requêtes. Pour cela, il est nécessaire de déléguer la tâche aux employés de Coexel. Cependant, leur outil ne leur permettait pas au moment de notre étude de créer différents bouquets de sources (auxquels sont appliquées différentes requêtes) au sein d’une même thématique.

Après avoir appris qu’ils ne pouvaient ni intégrer toutes nos sources (notamment les sources concernant les offres d’emploi et la source renseignant sur les brevets de semi-conducteurs aux USA), ni distinguer différents bouquets de sources au sein d’une même thématique, nous avons décidé de ne pas observer les résultats obtenus au moyen des sources manuelles, le paramétrage se différenciant trop des deux autres plateformes de veille et représentant un biais trop important pour l’analyse des résultats obtenus.

3.5 Collecte des données

3.5.1 Structure des dossiers de veille au sein des plateformes

Nous avons, au préalable, conçu la structure des dossiers et des sous-dossiers nécessaire à la collecte des informations en même temps que le paramétrage des plateformes.

En effet, après avoir analysé la manière dont étaient retournées les informations sur les différents outils, et anticipé nos besoins pour la phase de la récolte des résultats, nous avons créé sur chacune des plateformes :

- un dossier par axe de veille pour les “sources automatiques” (requêtes appliquées directement aux bases de données des plateformes),
- un dossier par bouquet de sources pour les “sources manuelles” (requêtes appliquées sur un bouquet de sources définies et importées manuellement). Plusieurs requêtes peuvent interroger un même bouquet de sources, et donc se trouver dans le même dossier.

En prévision de l’analyse des résultats retournés, pour chacun de ces dossiers, nous avons créé deux sous-dossiers (BRUIT et VALIDE), nous permettant par la suite de sauvegarder l’ensemble des informations collectées selon leur pertinence au regard de la requête et de nos

attentes. Après quelques manipulations, nous avons remarqué qu'en appliquant cette manipulation, certains résultats disparaissaient de la plateforme de Digimind (une information classée dans un axe, si elle existait aussi au sein d'un autre axe, s'effaçait). Pour éviter ce biais, et pour nous permettre plus de simplicité et de rapidité dans notre procédure de collecte des données, nous avons donc décidé de supprimer cette étape, la validité de l'information apparaissant déjà dans des tableurs externes aux plateformes créés à cet effet (cf. chapitre 3.5.4 *Saisie des résultats*).

3.5.2 Phase de veille

Initialement, le lancement de la phase de veille était planifié le 1er août 2017 pour une durée de 2 mois. En raison des difficultés rencontrées dans le paramétrage des plateformes de veille décrites dans le chapitre précédent, nous avons été contraintes de repousser d'une semaine la collecte des données. Nous avons ainsi réalisé notre veille sur les trois sujets et les trois plateformes évaluées du 7 août au 30 septembre 2017.

3.5.3 Choix des indicateurs

Pour répondre à l'objectif de notre étude qui est de définir quelle plateforme est la plus performante, nous avons défini les critères permettant de comparer les résultats retournés par chacune d'entre elles. Ces indicateurs ont été tirés des comparatifs de veille existants (cf. chapitre 2.2.2 *Les comparatifs d'outils de veille*) et ont également été définis en fonction des possibilités offertes par les différentes plateformes. D'autres indicateurs tels que l'analyse des doublons ou la fraîcheur des informations (c'est-à-dire la date de publication des informations sur le web) auraient également pu être définis, mais au vu du temps à disposition et pour des raisons de faisabilité, nous n'avons pas pu les développer pour notre étude.

Tableau 2 : Les indicateurs définis pour évaluer la qualité des plateformes

| INDICATEURS | EXPLICATIONS |
|-------------|--|
| Volumétrie | Nombre total de résultats collectés → But : identifier le nombre de résultats collectés par chacune des plateformes. |
| Pertinence | Adéquation entre les résultats attendus et les résultats reçus → But : identifier le nombre de résultats correspondant aux enjeux de l'axe de veille. |
| Précision | Rapport entre le nombre de résultats pertinents retournés et la volumétrie, exprimé en % → But : identifier le pourcentage de documents pertinents par rapport au nombre total de résultats retournés |

| | |
|--|---|
| <p>Complémentarité de l'ensemble des résultats et complémentarité des résultats pertinents</p> | <p>Rapport entre le nombre de résultats uniques d'une plateforme (dont les URLs n'ont pas été retournées par d'autres plateformes) et l'ensemble des résultats renvoyés par cette même plateforme, exprimé en %</p> <p>En ce qui concerne les requêtes dont les résultats ont été échantillonnés, la comparaison des résultats pertinents n'a pas été effectuée (il n'aurait pas été significatif de ne comparer que les résultats pertinents des échantillons, et impossible de traiter tous les résultats pour identifier l'intégralité des résultats pertinents de ces requêtes).</p> <p>→ But : identifier si les plateformes possèdent une plus-value en rapportant une part conséquente de résultats uniques.</p> |
| <p>Rapidité de capture des résultats et rapidité de capture des résultats pertinents</p> | <p>Date de collecte du résultat sur la plateforme</p> <p>→ But : identifier si une plateforme collecte un document avant une autre, lorsque le même résultat est retourné par différentes plateformes.</p> |

A propos de la pertinence, trois réponses pouvaient être données :

- VALIDE, lorsque l'information répond aux attentes du veilleur au regard de l'axe de veille concerné (et non en lien avec des requêtes spécifiques car certains dossiers au sein des plateformes comprennent les résultats de plusieurs requêtes),
- BRUIT, lorsque le veilleur estime que l'information ne répond pas aux enjeux de l'axe de veille,
- NON ANALYSABLE, lorsque, pour plusieurs raisons possibles, le veilleur ne peut pas juger de la pertinence de l'information (par ex., nécessité de créer un compte pour lire l'article, site malveillant qu'un anti-virus rend impossible à ouvrir, URL qui n'existe plus au moment de la consultation, etc.).

Comme décrit dans notre revue de littérature, le critère de pertinence est éminemment subjectif, car il dépend entièrement du jugement du veilleur. La simple présence des mots-clés dans la page retournée par la plateforme ne suffit pas à rendre l'information pertinente. En effet, un document pertinent est un document dans lequel l'utilisateur peut trouver l'information dont il a besoin (Nie 2016a). Dans le cadre de notre étude, il était nécessaire que l'information réponde aux enjeux de l'axe de veille, exprimés grâce aux requêtes. Bien que nous ayons pris soin de convenir ensemble du niveau d'exigence à adopter pour juger de la pertinence d'un résultat, ce point demeure un biais potentiel.

3.5.4 Saisie des résultats

Afin de répartir la charge de travail, plutôt que d'exporter l'ensemble des résultats à l'issue de la période de veille, nous avons pris le parti de collecter autant que faire se peut les résultats au fur et à mesure de leur apparition sur les différentes plateformes de veille. Pour cela, nous avons réalisé des tableurs permettant, pour chacune des requêtes réalisées dans chacune des plateformes, de relever :

- le titre du résultat (qui est celui apparaissant sur la plateforme),
- le nom du site web où apparaît le résultat collecté,
- l'URL du résultat,
- la date de l'importation du résultat sur la plateforme,
- la pertinence de l'information retournée.

Tableau 3 : Exemple de tableur de saisie des résultats

| Titre de l'article | Nom du site | URL | Date d'importati | Pertinence (BRUIT/VALIDE) |
|---|-------------|---|------------------|---------------------------|
| Vishay Intertechnology ENYCAP™ Electrical Double-Layer Energy Storage Capacitors Of | vishay | http://www.vishay | 16/08/2017 | BRUIT |
| Vishay Intertechnology Zhuhai, China Manufacturing Facility Earns ISO/TS 16949 Certific | vishay | http://www.vishay | 06/09/2017 | BRUIT |
| Vishay Intertechnology Launches Industry-First High Frequency IHLP® Inductor Series, B | vishay | http://www.vishay | 14/09/2017 | BRUIT |
| Director Merger and Acquisition (f/m) | infineon | https://www.infineon | 15/09/2017 | BRUIT |
| Assembly Manager Subcontractor Management (f/m) | infineon | https://www.infineon | 27/09/2017 | BRUIT |
| Vishay Intertechnology Wet Tantalum Capacitor Offers Industry-High Capacitance, Design | vishay | http://www.vishay | 30/09/2017 | BRUIT |
| Infineon strengthens its expertise in voice-controlled human machine interface with strateg | infineon | https://www.infineon | 06/09/2017 | VALIDE |
| STMicroelectronics Enters the CAC40 Paris Stock Index | st | http://investors.st | 08/09/2017 | VALIDE |

Tableur comportant les résultats liés à la requête SM_Axe2_R4 de la plateforme Sindup sur le sujet de veille Lemsys (semi-conducteur de puissance)

Certaines requêtes ont très rapidement rapporté trop de résultats, rendant impossible un traitement manuel au fur et à mesure de la phase de veille. Nous avons donc dû les exporter au terme de la collecte afin de procéder à un échantillonnage. Nous avons paramétré au mieux les plateformes afin que les mêmes informations que celles saisies manuellement soient recueillies. Nous avons cependant rencontré des difficultés pour obtenir les dates d'importation des résultats dans la plateforme Digimind lors de leur export. Pour cette raison, nous n'avons pas pu comparer les dates de collecte des résultats de la requête en Sciences de l'information SM_Axe1_R1.

3.5.5 Echantillonnage

Comme expliqué ci-dessus, au terme des 10 premiers jours de la collecte, nous avons relevé que certaines requêtes rapportaient un nombre de résultats trop important pour qu'il soit possible de les consulter et de les analyser intégralement. C'est ainsi qu'il a été décidé que les résultats de certaines requêtes ne seraient exportés et traités qu'à l'issue de la période de veille. En effet, ce n'est que lorsque le nombre total de résultats est connu qu'il est possible de procéder à un échantillonnage.

Pour échantillonner nos données, qui sont de type quantitatif, nous avons décidé de procéder à un échantillonnage aléatoire (par randomisation, ou sélection aléatoire par ordinateur) car il permet d'obtenir un résultat stable.

Tableau 4 : Méthode d'échantillonnage par sélection aléatoire par ordinateur

| | A | B | C | D | E | F | G | H |
|---|------|------------|---------------------------------------|---|---|------------|-----------|------|
| 1 | | | | | | | | |
| 2 | Cle | | | | | | | |
| 3 | | 179 | | | | | | |
| 4 | | | | | | | | |
| 5 | Cle | Cle_Random | Titre | Nom du site | URL | Date | Pertinenc | lien |
| 6 | 3831 | 617 | Why the World's CEOs Fear Paul Singer | http://bloomberg.com | https://www.bloomberg.com/graphics/2017-elliott-management/ | 07.08.2017 | | lien |
| 7 | | | | | | | | |

Une formule insérée dans le logiciel Excel permet d'obtenir de manière aléatoire des numéros représentant des résultats à analyser. Les données sont automatiquement complétées, le veilleur doit simplement cliquer sur le lien internet mis à disposition afin de juger de la pertinence du résultat, puis compléter la colonne Pertinence selon les 3 catégories possibles (VALIDE, BRUIT, NON ANALYSABLE).

Le risque d'erreur est alors lié à la taille de l'échantillon, qui doit être définie selon le risque d'erreur accepté pour l'étude. Plusieurs critères entrent en jeu lors de l'élaboration de la taille de l'échantillon :

- l'objectif de la recherche,
- le budget et le temps,
- le taux d'erreur acceptable,
- le niveau de variation estimé de la population (ou hétérogénéité) (Ruch 2016).

Ainsi, si la marge d'erreur est réduite avec un échantillon de taille importante, il faut aussi prendre en compte la faisabilité de la démarche, le temps imparti à ce projet étant limité.

Afin que nos échantillons soient représentatifs de l'hétérogénéité de nos résultats, nous avons réalisé les différents échantillonnages en calculant progressivement la précision du nombre de résultats valides retournés, selon le calcul suivant :

$$\text{Précision } P = \text{TP}/(\text{TP}+\text{FP})$$

où TP = valide

et FP = bruit

En effet, ce calcul permet de donner en pourcentage le nombre de documents retrouvés qui sont corrects (Gobeill 2016).

Par la mise en place d'un graphique généré automatiquement, nous avons pu identifier le moment où le taux de précision se stabilise, et où, donc, la collecte de résultats supplémentaires devient inutile.

Tableau 5 : Calcul automatisé de la précision

| Numero | Cle | NB Valide | %Valide | Variation |
|--------|-----|-----------|---------|-----------|
| 1 | 1 | 1 | 100% | |
| 2 | 2 | 2 | 100% | 0% |
| 3 | 3 | 3 | 100% | 0% |
| 5 | 5 | 4 | 80% | 5% |
| 6 | 6 | 5 | 83% | 3% |
| 7 | 7 | 6 | 86% | 2% |
| 8 | 8 | 7 | 88% | 2% |
| 9 | 9 | 8 | 89% | 1% |
| 10 | 10 | 9 | 90% | 1% |
| 11 | 11 | 10 | 91% | 1% |
| 12 | 12 | 11 | 92% | 1% |
| 13 | 13 | 12 | 92% | 1% |
| 14 | 14 | 13 | 93% | 1% |
| 17 | 17 | 14 | 82% | 1% |

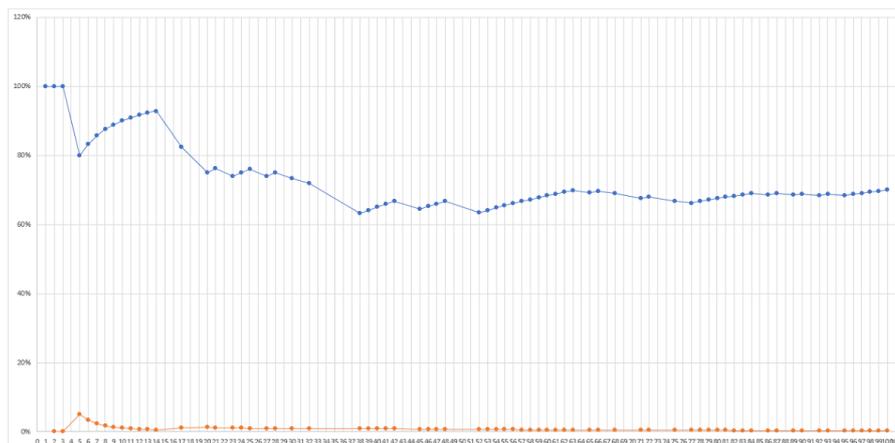
A mesure que le premier tableau est complété par le veilleur, un deuxième tableau généré automatiquement permet de mesurer l'évolution de la précision pour la requête concernée.

Ci-dessus, nous pouvons observer que les 3 premiers résultats sont valides, le pourcentage de résultats valides s'élève donc à 100%. Le 4ème résultat n'est pas pertinent, ce qui fait diminuer le pourcentage de résultats valides à 80%, et modifie la variation à 5%.

Ce tableau correspond au calcul de la précision $P : \text{TP}/(\text{TP}+\text{FP})$, où TP représente les résultats valides et FP le bruit.

Tableau 6 : Evolution de la pertinence des résultats

| Numero | Cle | NB Valide | %Valide | Variation |
|--------|-----|-----------|---------|-----------|
| 1 | 1 | 1 | 100% | |
| 2 | 2 | 2 | 100% | 0% |
| 3 | 3 | 3 | 100% | 0% |
| 5 | 5 | 4 | 80% | 5% |
| 6 | 6 | 5 | 83% | 3% |
| 7 | 7 | 6 | 86% | 2% |
| 8 | 8 | 7 | 88% | 2% |
| 9 | 9 | 8 | 89% | 1% |
| 10 | 10 | 9 | 90% | 1% |
| 11 | 11 | 10 | 91% | 1% |
| 12 | 12 | 11 | 92% | 1% |
| 13 | 13 | 12 | 92% | 1% |
| 14 | 14 | 13 | 93% | 1% |
| 17 | 17 | 14 | 82% | 1% |
| 20 | 20 | 15 | 75% | 1% |
| 21 | 21 | 16 | 76% | 1% |
| 23 | 23 | 17 | 74% | 1% |
| 24 | 24 | 18 | 75% | 1% |
| 25 | 25 | 19 | 76% | 1% |
| 27 | 27 | 20 | 74% | 1% |
| 28 | 28 | 21 | 75% | 1% |
| 30 | 30 | 22 | 73% | 1% |
| 32 | 32 | 23 | 72% | 1% |
| 38 | 38 | 24 | 63% | 1% |
| 39 | 39 | 25 | 64% | 1% |
| 40 | 40 | 26 | 65% | 1% |
| 41 | 41 | 27 | 66% | 1% |
| 42 | 42 | 28 | 67% | 1% |
| 45 | 45 | 29 | 64% | 1% |
| 46 | 46 | 30 | 65% | 1% |
| 47 | 47 | 31 | 66% | 1% |
| 48 | 48 | 32 | 67% | 1% |
| 52 | 52 | 33 | 63% | 1% |
| 53 | 53 | 34 | 64% | 1% |
| 54 | 54 | 35 | 65% | 1% |
| 55 | 55 | 36 | 65% | 1% |



Ce 2ème tableau génère automatiquement un graphique qui permet de visualiser l'évolution de la pertinence des résultats retournés. Lorsque la courbe s'est suffisamment stabilisée, nous estimons avoir atteint ou tout du moins approché le point de saturation, et cessons la procédure d'échantillonnage. Le taux de précision de l'échantillon est alors représentatif du taux de précision que nous aurions obtenu si nous avions analysé l'ensemble des résultats récoltés par la requête.

Nous pouvons alors évoquer, tout comme pour l'échantillonnage de données qualitatives, une recherche du point de saturation (moment où l'on peut considérer que l'on n'obtiendra plus de nouvel input) (Ruch 2016). Au regard du nombre de résultats très important renvoyé par certaines requêtes, cette technique nous permet d'effectuer ainsi un échantillonnage raisonné.

3.5.6 Traitement des données

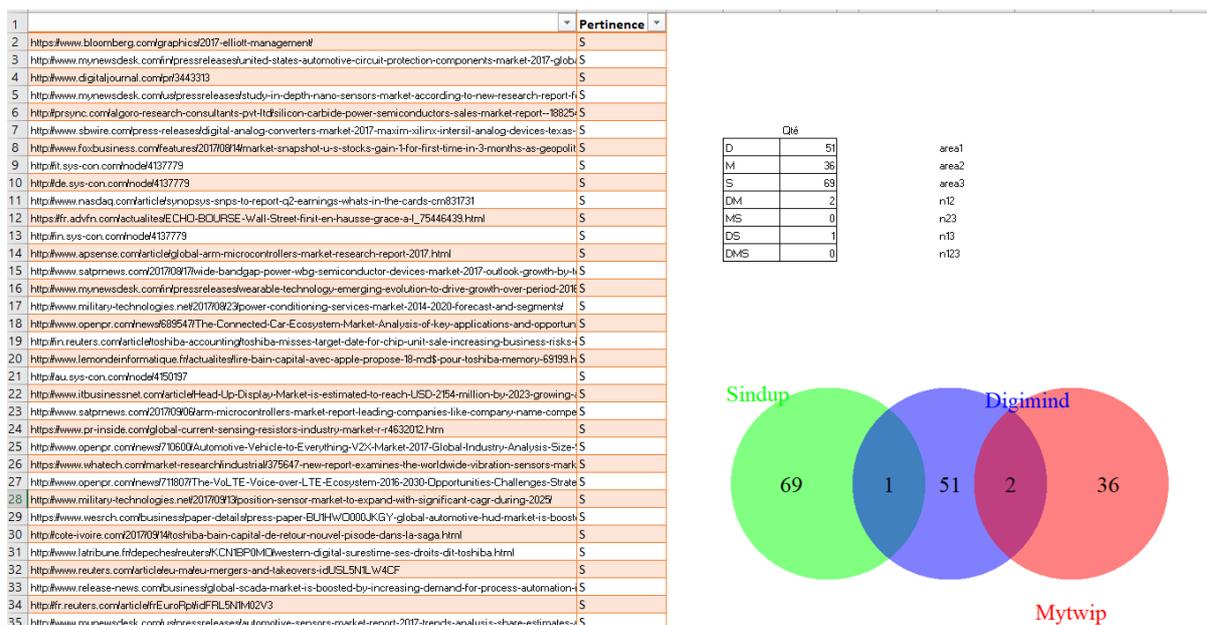
A l'aide des données récoltées dans les différents tableurs, nous avons, pour chaque requête sur chacune des plateformes, calculé les premiers indicateurs que nous avons définis : la volumétrie et la pertinence des résultats ainsi que la précision des plateformes qui en découle.

Ces documents ont été ensuite utilisés comme sources de données pour la création de documents de synthèse via le logiciel Microsoft Query.

Pour chacune des requêtes, ce logiciel a donc permis de croiser les données obtenues par les différentes plateformes, de manière à pouvoir identifier de manière automatique :

- les résultats pertinents communs à 2 ou aux 3 plateformes (complémentarité des résultats obtenus),
- la date de collecte des résultats renvoyés par les différentes plateformes (rapidité de capture).

Tableau 7 : Croisement des résultats pertinents entre les 3 plateformes



Exemple de tableau collectant automatiquement les résultats pertinents des 3 plateformes pour la requête SA_Axe2_R3 (sujet Lemsys (semi-conducteur de puissance))

Tableau 8 : Croisement des dates de collecte entre les 3 plateformes

| URL | Digimind.Date | Mytwip.Date | Sindup.Date | Lien |
|--|---------------|-------------|-------------|------|
| http://www.boursorama.com/actualites/toshiba-pres-de-conclure-la-vente-de-ses-puces-a-bain-ff | 21.09.2017 | | 20.09.2017 | lien |
| http://www.lemondeinformatique.fr/actualites/lire-recap-it-malware-dans-cleaner-meilleur-dev | 22.09.2017 | 22.09.2017 | | lien |
| https://www.sciencesetavenir.fr/high-tech/bain-s-allie-a-des-groupes-high-tech-pour-les-puces-de | 16.09.2017 | 16.09.2017 | | lien |

Exemple de tableau comparant automatiquement les dates de collecte des résultats pertinents des 3 plateformes pour la requête SA_Axe2_R3 (sujet Lemsys (semi-conducteur de puissance))

Nous avons inscrit l'ensemble des résultats dans de nouveaux tableaux regroupant l'ensemble des requêtes liées à un sujet de veille et à une plateforme. Par ailleurs, lorsqu'un échantillonnage a dû être réalisé, nous avons également calculé le pourcentage de ce que cela représente par rapport au nombre de résultats total reçus.

Tableau 9 : Récapitulatif des résultats par plateforme et sujet de veille

| | Nombre de résultats | | Taille de l'échantillon | | % de l'échantillon | | Nb de résultats | | Total des résultats | | Nb de résultats pertinents | | % de résultats pertinents | | Nb de résultats | | % de résultats | |
|----------------------|---------------------|-------------|-------------------------|-----------------|--------------------|---------------|-----------------|----------------|---------------------|--|----------------------------|--|---------------------------|--|-----------------|--|----------------|--|
| | ou VOLUMETRIE | échantillon | /pop. totale | non analysables | analysables | ou PERTINENCE | ou PRECISION | non pertinents | non pertinents | | | | | | | | | |
| SM_Axe1_R1 | 411 | 111 | 27.0% | 9 | 102 | 47 | 46.1% | 55 | 53.9% | | | | | | | | | |
| SM_Axe1_R2 | 8 | / | | 0 | 8 | 5 | 62.5% | 3 | 37.5% | | | | | | | | | |
| SM_Axe1_R3 | 2 | / | | 0 | 2 | 2 | 100.0% | 0 | 0.0% | | | | | | | | | |
| SM_Axe1_R4 | 9 | / | | 0 | 9 | 5 | 55.6% | 4 | 44.4% | | | | | | | | | |
| SM_Axe1_R5 | 44 | / | | 1 | 43 | 28 | 65.1% | 15 | 34.9% | | | | | | | | | |
| SM_Axe1_R6 | 28 | / | | 0 | 28 | 22 | 78.6% | 6 | 21.4% | | | | | | | | | |
| SM_Axe1_R7 | 25 | / | | 0 | 25 | 19 | 76.0% | 6 | 24.0% | | | | | | | | | |
| SA_Axe1_R1 | 67 | / | | 0 | 67 | 4 | 6.0% | 63 | 94.0% | | | | | | | | | |
| SA_Axe1_R2 | 17 | / | | 0 | 17 | 1 | 5.9% | 16 | 94.1% | | | | | | | | | |
| SA_Axe1_R3 | 8 | / | | 0 | 8 | 1 | 12.5% | 7 | 87.5% | | | | | | | | | |
| SA_Axe1_R4 | 4 | / | | 0 | 4 | 1 | 25.0% | 3 | 75.0% | | | | | | | | | |
| SA_Axe1_R5 | 5 | / | | 0 | 5 | 2 | 40.0% | 3 | 60.0% | | | | | | | | | |
| SA_Axe1_R6 | 7 | / | | 0 | 7 | 2 | 28.6% | 5 | 71.4% | | | | | | | | | |
| SA_Axe1_R7 | 1 | / | | 1 | 0 | 0 | | 0 | | | | | | | | | | |
| SM_Axe2_R8 | 299 | 101 | | 2 | 99 | 22 | 22.2% | 77 | 77.8% | | | | | | | | | |
| SM_Axe2_R9 | 51 | / | | 0 | 51 | 40 | 78.4% | 11 | 21.6% | | | | | | | | | |
| SM_Axe2_R10 | 18 | / | | 0 | 18 | 9 | 50.0% | 9 | 50.0% | | | | | | | | | |
| SA_Axe2_R8 | 275 | 152 | 55.3% | 2 | 150 | 32 | 21.3% | 118 | 78.7% | | | | | | | | | |
| SA_Axe2_R9 | 210 | 77 | 36.7% | 3 | 74 | 0 | 0.0% | 74 | 100.0% | | | | | | | | | |
| Sources automatiques | 594 | 229 | 38.6% | 6 | 332 | 43 | 13.0% | 289 | 87.0% | | | | | | | | | |
| Sources manuelles | 895 | 212 | 23.7% | 12 | 385 | 199 | 51.7% | 186 | 48.3% | | | | | | | | | |
| TOTAL | 1489 | 441 | 29.6% | 18 | 717 | 242 | 33.8% | 475 | 66.2% | | | | | | | | | |

Tableau regroupant les résultats de la veille au sujet des Sciences de l'information effectuée sur la plateforme Sindup.

3.6 Analyse des résultats

Afin de permettre l'analyse de nos différents résultats, nous avons dans un premier temps réalisé 3 rapports, par sujet de veille (cf. Annexe 6).

Ces rapports ont permis de relever pour chacun des 3 sujets de veille :

- l'ensemble des résultats pour chacune des plateformes (avec calculs des taux de précision et visualisation de la volumétrie),
- les résultats par requête (avec calculs des taux de précision, volumétrie, complémentarité des résultats et enfin rapidité de capture),
- les résultats pour les sources manuelles (ou SM) (pour l'ensemble des requêtes le calcul de la volumétrie, de la précision des plateformes, de la complémentarité des résultats retournés et de la rapidité de capture),
- les résultats pour les sources automatiques (ou SA).

4. Présentation et analyse des résultats

Nous avons pu constater que les résultats obtenus sont très variés selon les sujets de veille, les approches (sources manuelles vs sources automatiques), les requêtes ou encore les indicateurs sélectionnés.

Nous dégageons ici les grandes tendances observées selon les cinq indicateurs que nous avons choisis pour définir la qualité d'une plateforme, l'ensemble de nos observations étant consignés dans les rapports Analyse SI, Analyse IoT et Analyse Lemsys (semi-conducteur de puissance) à disposition en annexes (cf. Annexe 6).

4.1 Volumétrie

Comme indiqué dans le chapitre développant notre méthodologie, la volumétrie est l'indicateur qui correspond au nombre total de résultats collectés. Premier indicateur relevé, il est à la base de nombreux autres calculs permettant des analyses plus approfondies.

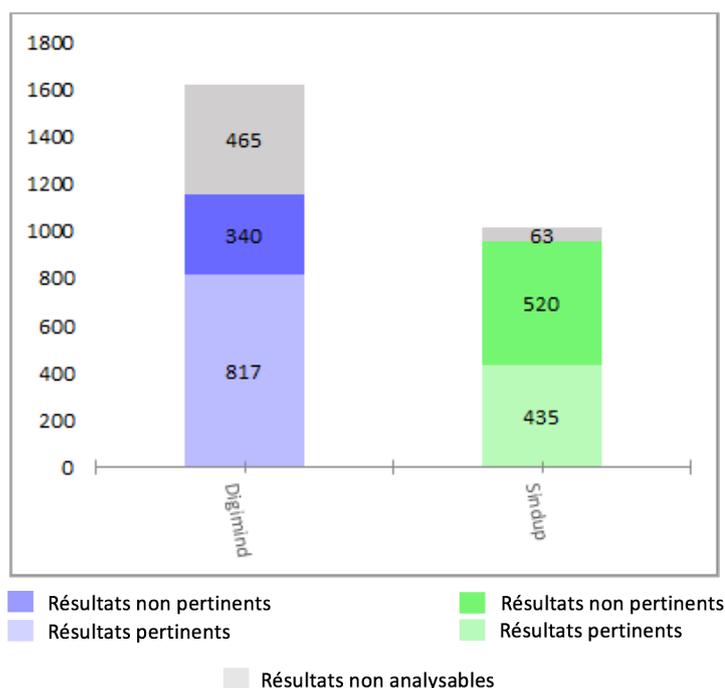
Pour rappel, trois types de résultats ont été relevés :

- les résultats pertinents (qui répondent au besoin documentaire),
- les résultats non pertinents (qui ne répondent pas au besoin documentaire),
- les résultats non analysables (qui sont impossible à consulter pour diverses raisons).

4.1.1 Volumétrie des sources manuelles (ou SM)

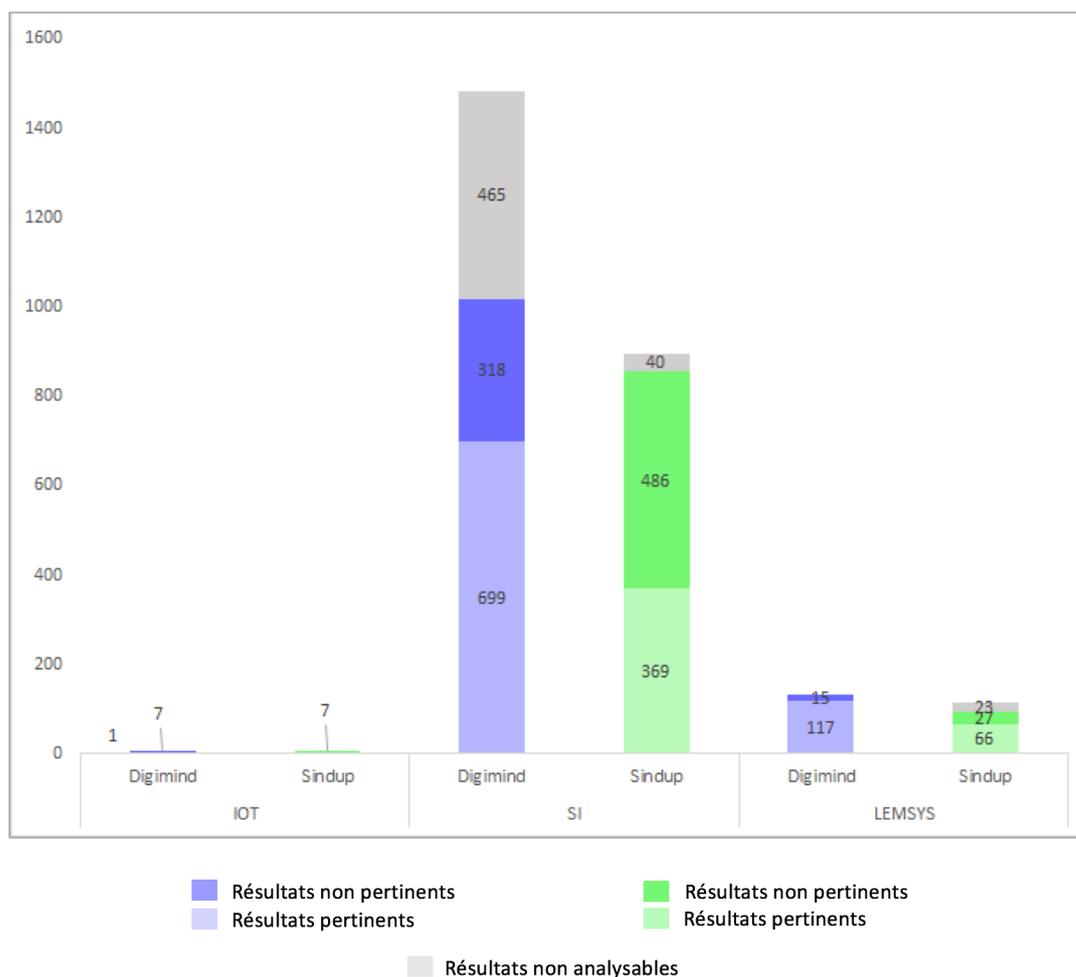
Comme cela a pu être expliqué plus haut, la plateforme MyTwip n'a pas pu être paramétrée selon cette approche. Les sources manuelles ne permettent donc de comparer que les plateformes Digimind et Sindup.

Figure 5 : SM : Nombre et type de résultats par plateforme



Si l'on observe l'ensemble des résultats renvoyés par les deux plateformes, tous sujets de veille confondus, on constate alors que la plateforme Digimind est celle qui possède la volumétrie la plus importante avec 1'622 résultats rapportés contre 1'018 pour Sindup.

Figure 6 : SM : Nombre et type de résultats par plateforme et par sujet de veille



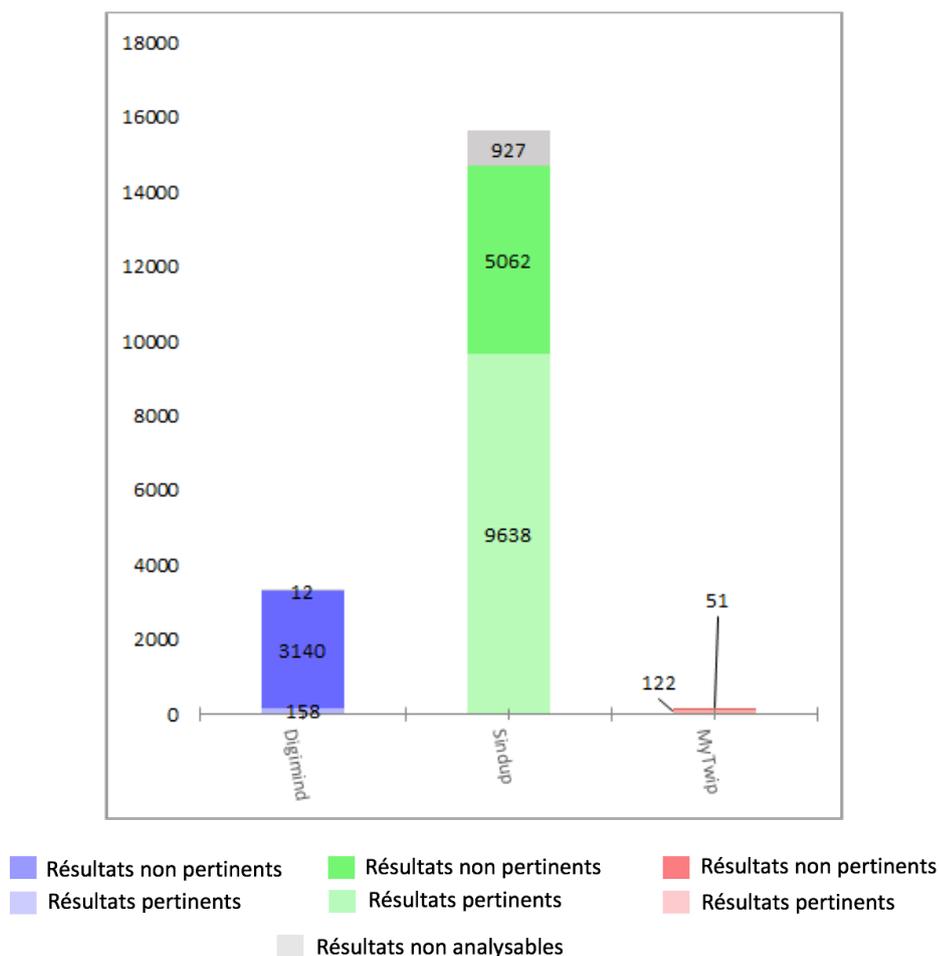
Cette tendance se confirme lorsque l'on observe chacun des trois sujets de veille.

On constate que certains sujets rapportent plus de résultats que d'autres, et ce de manière identique pour les deux plateformes. En effet, si de nombreux résultats ont été renvoyés pour le sujet sur les Sciences de l'information (1'482 résultats pour Digimind et 895 pour Sindup), les plateformes ont en revanche rapporté très peu de résultats concernant le sujet sur l'IoT (8 résultats pour Digimind, 7 pour Sindup).

Ce phénomène s'explique très certainement par le choix des sources qui ont été insérées manuellement (sites internet des industries pharmaceutiques, hôpitaux, associations de médecins et écoles suisses) qui ont été très peu actualisées durant la période de veille.

4.1.2 Volumétrie des sources automatiques (ou SA)

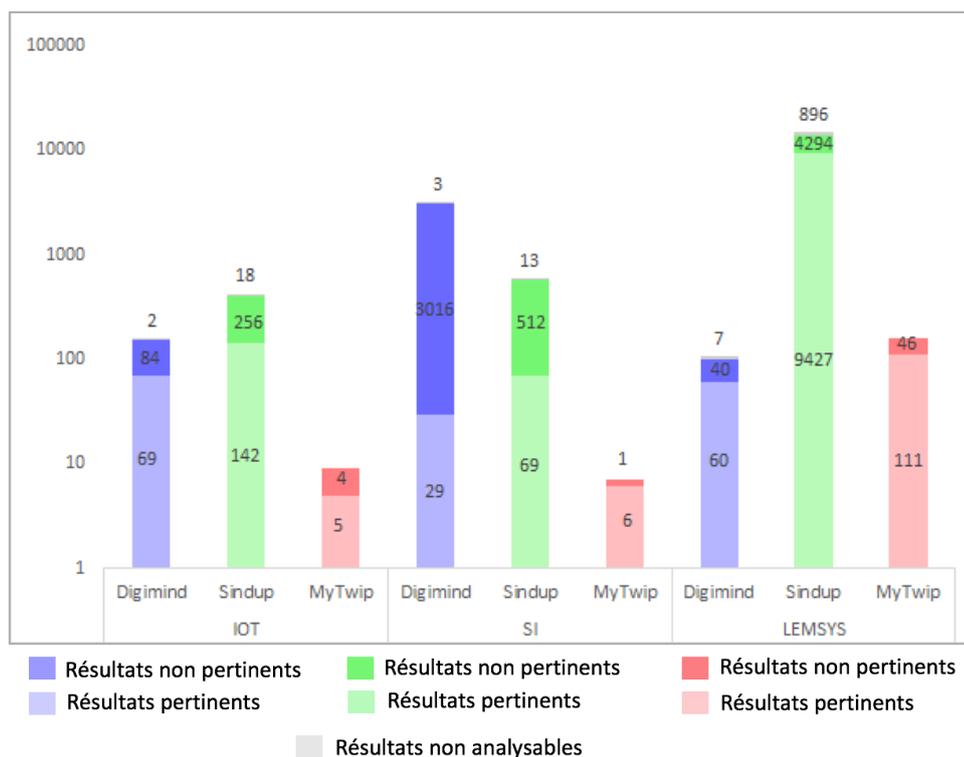
Figure 7 : SA : Nombre et type de résultats par plateforme



Si l'on s'attache à présent à observer la volumétrie des résultats renvoyés par le biais des sources automatiques (ou SA), alors on constate une très grande disparité entre les trois plateformes. Sindup est celle qui a renvoyé le plus grand nombre de résultats (15'627), suivie par Digimind (3'310), puis pour finir MyTwip (avec seulement 173 résultats).

En effet, deux requêtes à propos de Lemsys (semi-conducteur de puissance) ont ramené énormément de résultats sur la plateforme Sindup (SA_Axe2_R3 : 5'939 résultats et SA_Axe3_R5 : 8'641 résultats). Egalement, deux requêtes pour les Sciences de l'information ont rapporté énormément de bruit sur la plateforme Digimind pour des raisons non identifiées (SA_Axe1_R5 : 1'819 résultats, SA_Axe1_R6 : 1'038 résultats).

Figure 8 : SA : Nombre et type de résultats par plateforme et par sujet de veille



Au regard des deux autres plateformes, on peut observer que MyTwip est une plateforme qui renvoie très peu de résultats, ce qui se vérifie pour 2 des 3 sujets de veille :

IoT : Digimind : 155 ; Sindup : 416 ; MyTwip : 9 ;

SI : Digimind : 3'048 ; Sindup : 594 ; MyTwip : 7

En revanche, concernant Lemsys (semi-conducteur de puissance), c'est alors la plateforme Digimind qui a rapporté le moins de résultats (107 contre 157 pour MyTwip et 14'617 pour Sindup).

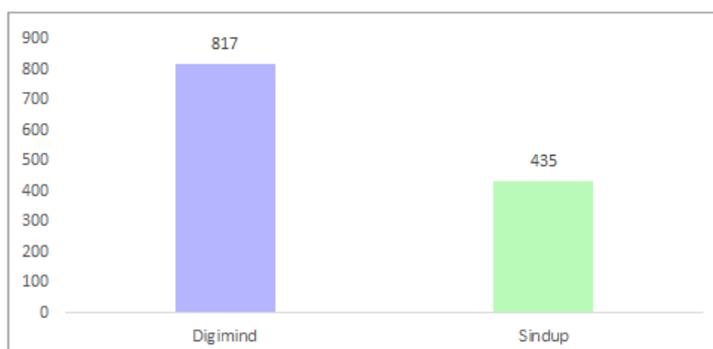
4.2 Pertinence

Deuxième indicateur de notre étude, la pertinence correspond à l'adéquation entre les résultats attendus et les résultats reçus, l'objectif étant d'identifier le nombre de résultats correspondant aux enjeux de l'axe de veille.

Cet indicateur précise la volumétrie en indiquant le nombre de résultats pertinents renvoyés par les plateformes.

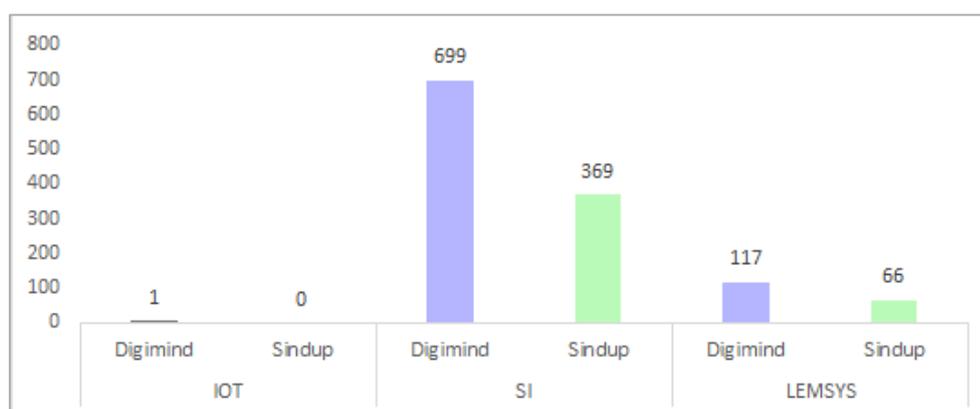
4.2.1 Pertinence des sources manuelles (ou SM)

Figure 9 : SM : Nombre de résultats pertinents par plateforme



Concernant les sources manuelles, si Digimind est la plateforme qui possède la volumétrie la plus importante tous résultats confondus, elle est aussi celle qui renvoie le plus de résultats pertinents, avec 817 résultats contre 435 pour Sindup.

Figure 10 : SM : Nombre de résultats pertinents par plateforme et par sujet de veille

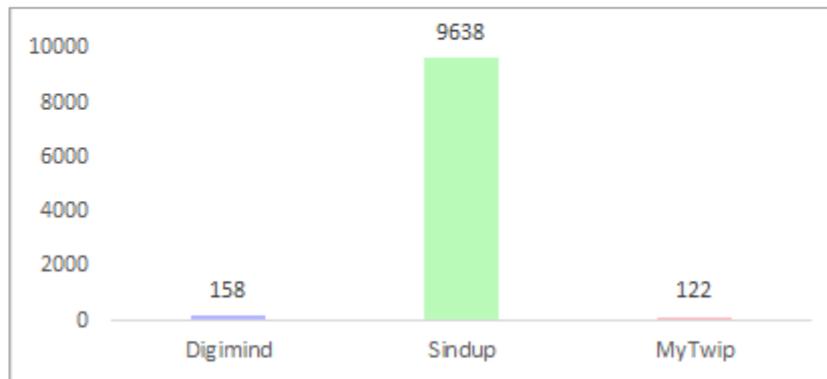


Cette tendance se confirme également lorsque l'on observe chacun des 3 sujets de veille, avec des écarts plus ou moins creusés.

4.2.2 Pertinence des sources automatiques (ou SA)

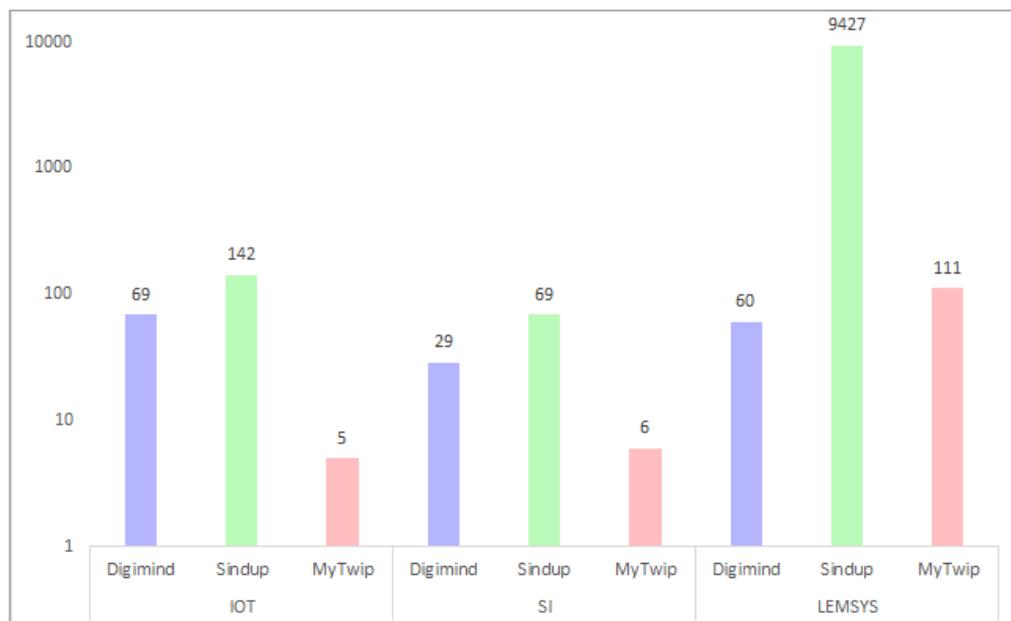
Pour les 4 requêtes évoquées plus haut ayant rapporté énormément de résultats, des échantillonnages ont dû être effectués afin de pouvoir estimer le nombre de résultats pertinents, non pertinents et non analysables.

Figure 11 : SA : Nombre de résultats pertinents par plateforme



Ainsi, selon les statistiques permises par ces échantillonnages, il est possible d'affirmer que si Sindup est la plateforme qui a renvoyé le plus de résultats, elle est aussi celle qui a rapporté le plus de résultats pertinents (9'638), suivie de loin par Digimind (158), puis par MyTwip (122). Ces deux dernières plateformes produisent donc du silence, au regard des résultats renvoyés par Sindup.

Figure 12 : SA : Nombre de résultats par plateforme et par sujet de veille



Sindup est la plateforme qui rapporte le plus de résultats pertinents pour chacun des sujets de veille, avec un écart particulièrement marqué concernant le sujet Lemsys (semi-conducteur de puissance).

Si Digimind rapporte plus de résultats pertinents que MyTwip pour les sujets sur l'IoT (69 résultats contre 5) et sur les Sciences de l'information (29 résultats contre 6), c'est en revanche MyTwip qui se place en deuxième position pour le sujet Lemsys (semi-conducteur de puissance) avec 111 résultats pertinents, contre 60 pour Digimind.

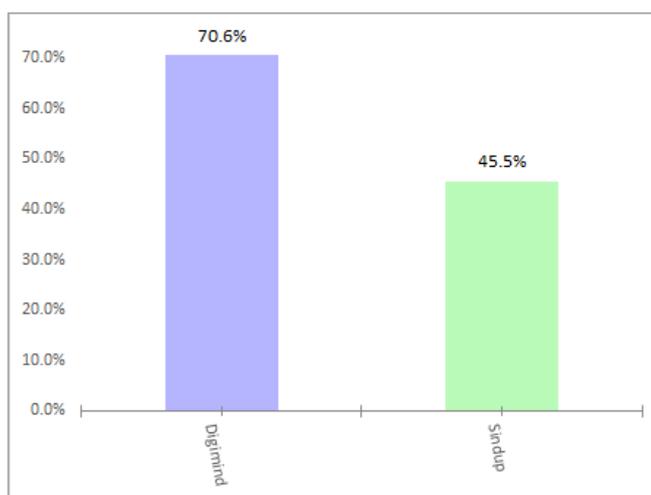
4.3 Précision

A différencier de l'indicateur précédent, la précision indique le rapport entre le nombre de résultats pertinents retournés et la volumétrie, et est exprimé en pourcentage.

La précision permet d'identifier le pourcentage de documents pertinents par rapport au nombre total de résultats analysables retournés. Nous n'avons en effet pas pris en considération les résultats non analysables dans nos calculs pour cet indicateur.

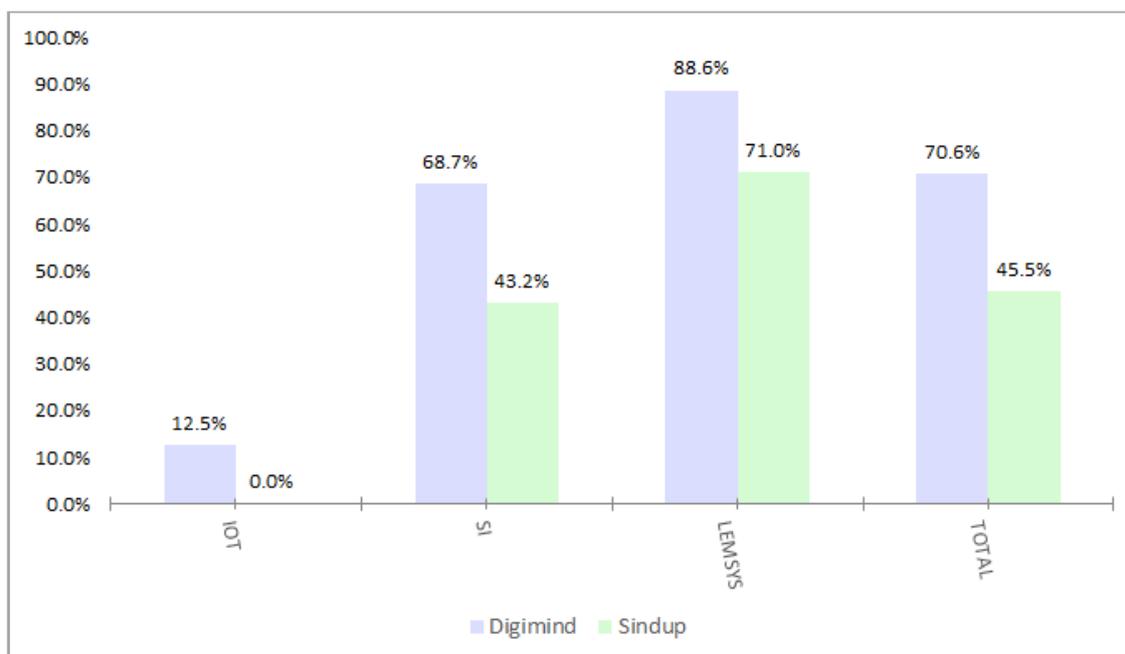
4.3.1 La précision des sources manuelles (ou SM)

Figure 13 : SM : Précision par plateforme



Tous sujets de veille confondus, la précision de la plateforme Digimind est largement supérieure à celle de sa concurrente (70.6% contre 45.5%), comme l'illustre le graphique ci-dessus.

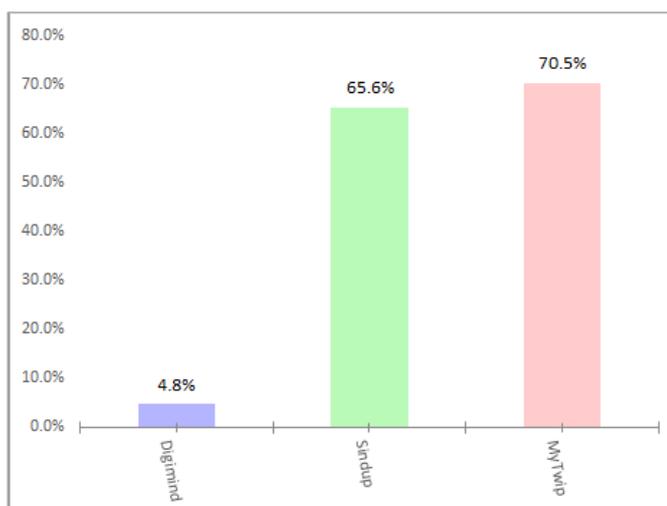
Figure 14 : SM : Précision par sujet et par plateforme



Cette tendance se confirme à l'analyse de chacun des trois sujets de veille.

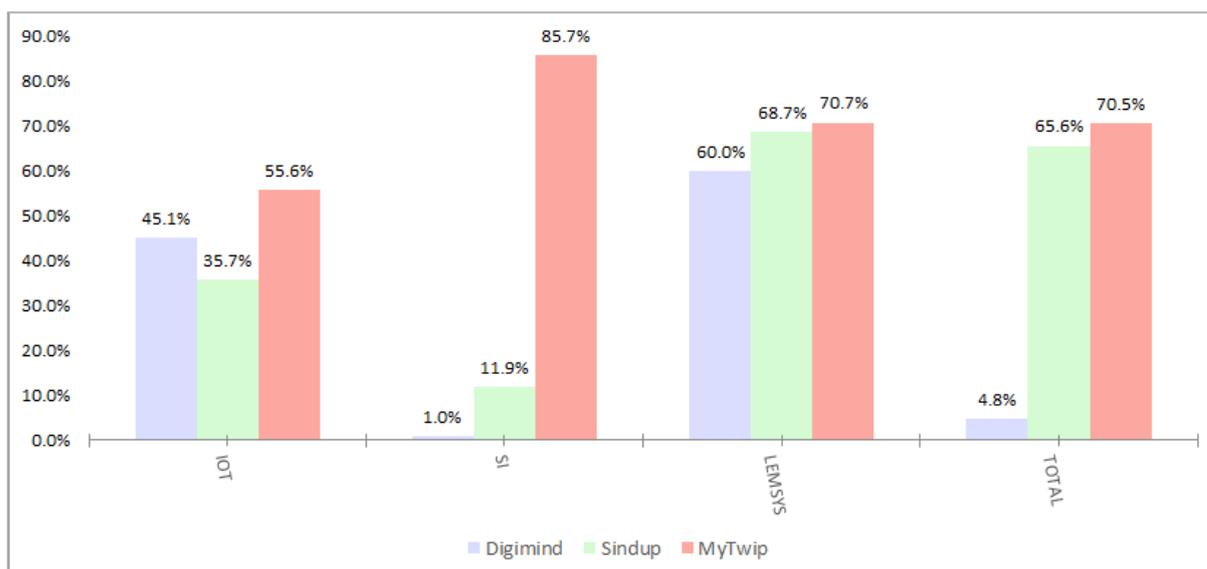
4.3.2 La précision des sources automatiques (ou SA)

Figure 15 : SA : Précision par plateforme



L'analyse est tout à fait différente lorsque l'on observe les résultats issus des sources automatiques. En effet, à cause des deux requêtes en Sciences de l'information citées plus haut qui ont rapporté énormément de bruit, la précision de Digimind chute à 4.8%, la plateforme se retrouvant alors en 3ème position derrière MyTwip avec 70.5% et Sindup avec 65.6%.

Figure 16 : SA : Précision par sujet et par plateforme



A l'analyse de chacun des sujets de veille, MyTwip se révèle être à chaque fois la plateforme possédant le meilleur taux de précision, avec un écart particulièrement marqué pour le sujet traitant des Sciences de l'information (avec 85.7% de précision, contre 11.9% pour Sindup et 1% pour Digimind). Digimind est la plateforme la moins précise lorsqu'il s'agit des sujets traitant des Sciences de l'information et de Lemsys (semi-conducteur de puissance). Concernant l'IoT, c'est cependant Sindup qui se trouve être la plateforme possédant le plus bas taux de précision (35.7%, contre 55.6% pour MyTwip et 45.1% pour Digimind).

4.4 Complémentarité

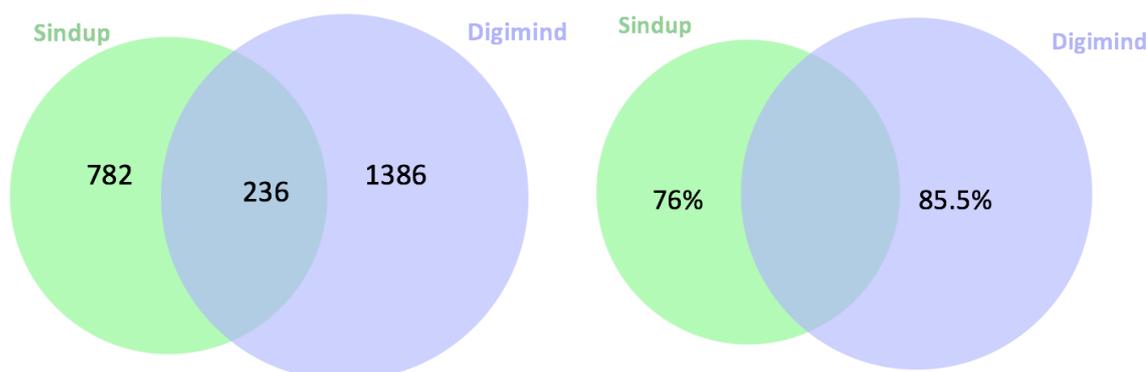
Comme expliqué dans le chapitre présentant notre méthodologie, la complémentarité représente le rapport entre le nombre de résultats uniques d'une plateforme (dont les URLs n'ont pas été retournées par d'autres plateformes) et l'ensemble des résultats renvoyés par cette même plateforme, exprimé en %.

Nous nous sommes intéressées, d'une part, à la complémentarité de l'ensemble des résultats collectés, et d'autre part, uniquement à la complémentarité des résultats pertinents collectés. Un grand nombre de requêtes ayant été échantillonnées, la comparaison des résultats pertinents n'a parfois pas pu être faite car il n'aurait pas été significatif de ne comparer que les résultats pertinents des échantillons.

Ci-après, nous présentons donc la complémentarité de l'ensemble des résultats collectés pour les sources manuelles (SM) et pour les sources automatiques (SA).

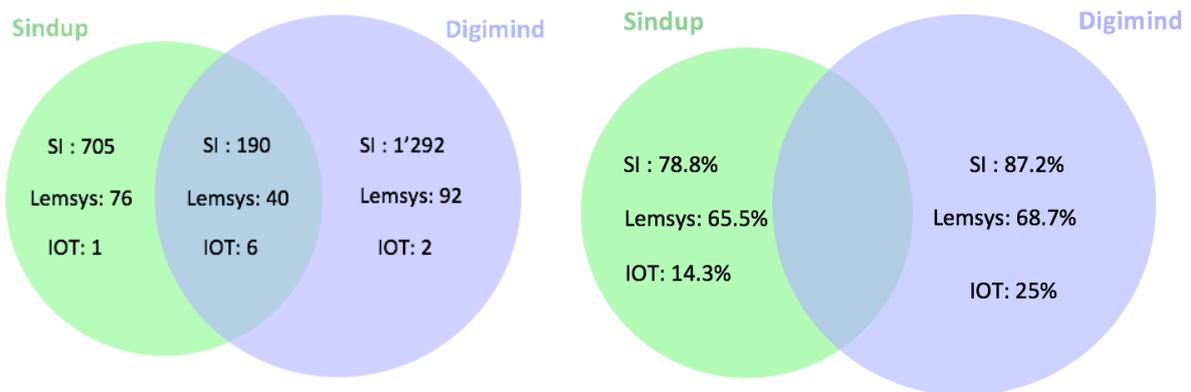
4.4.1 La complémentarité des sources manuelles (ou SM)

Figure 17 : SM : Complémentarité des résultats retournés



Les plateformes Digimind et Sindup ont retourné 236 résultats en commun sur l'ensemble des données collectées via les sources manuelles (SM). Digimind possède 1'386 résultats uniques (qui n'ont pas été également renvoyés par sa concurrente Sindup) et Sindup 782. Cela nous permet de calculer leur taux de complémentarité qui pour la première est de 85.5% et pour la seconde, de 76%. Digimind rapporte donc davantage de résultats qui lui sont propres.

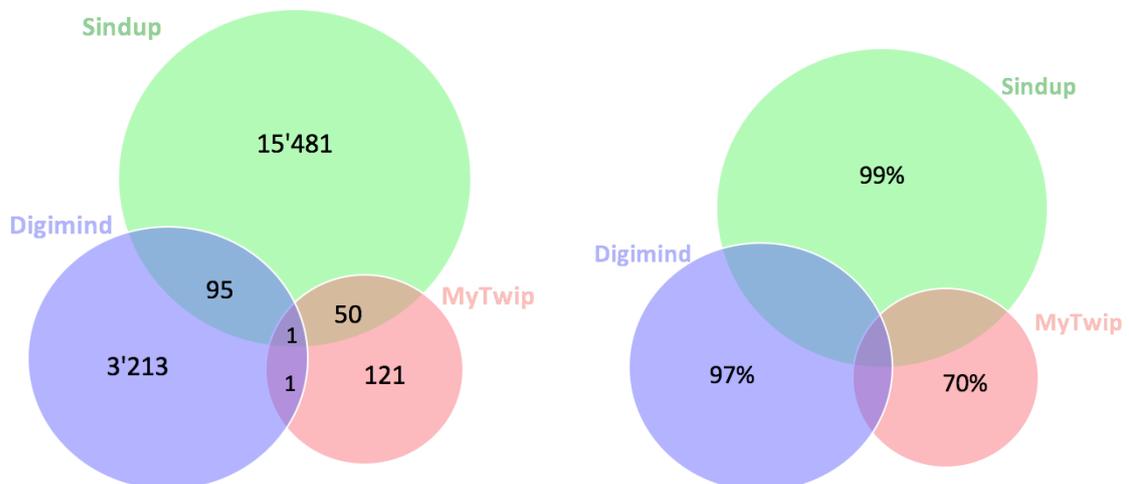
Figure 18 : SM : Complémentarité des résultats retournés par sujets



En analysant le même indicateur selon chacun des sujets de veille, nous observons que plus les plateformes ont collecté de résultats et plus celles-ci en possèdent un nombre important en commun. Ainsi, les plateformes ont retourné 190 résultats en commun pour le sujet SI, 40 pour Lemsys (semi-conducteur de puissance) et seulement 6 pour l'IoT. Par ailleurs, nous observons que pour chacun des sujets de veille, il s'agit de la plateforme Digimind qui possède le taux de complémentarité le plus élevé. Nous en déduisons que celle-ci possède donc davantage de résultats uniques que sa concurrente Sindup. Notons tout de même que le taux de complémentarité est très variable d'un sujet de veille à l'autre. En effet, en ce qui concerne l'IoT, très peu de résultats sont uniques à chacune des plateformes. Elles ont donc un taux de complémentarité très faible (14.3% pour Sindup et 25% pour Digimind). En revanche, sur le sujet Lemsys (semi-conducteur de puissance), ce taux est un peu plus élevé car les deux plateformes ont renvoyé moins de résultats communs (proportionnellement) : 65.5% pour Sindup et 68.7% pour Digimind. C'est sur le sujet SI que les plateformes ont renvoyé le plus de résultats uniques avec près de 80% pour Sindup et près de 88% pour Digimind.

4.4.2 La complémentarité des sources automatiques (ou SA)

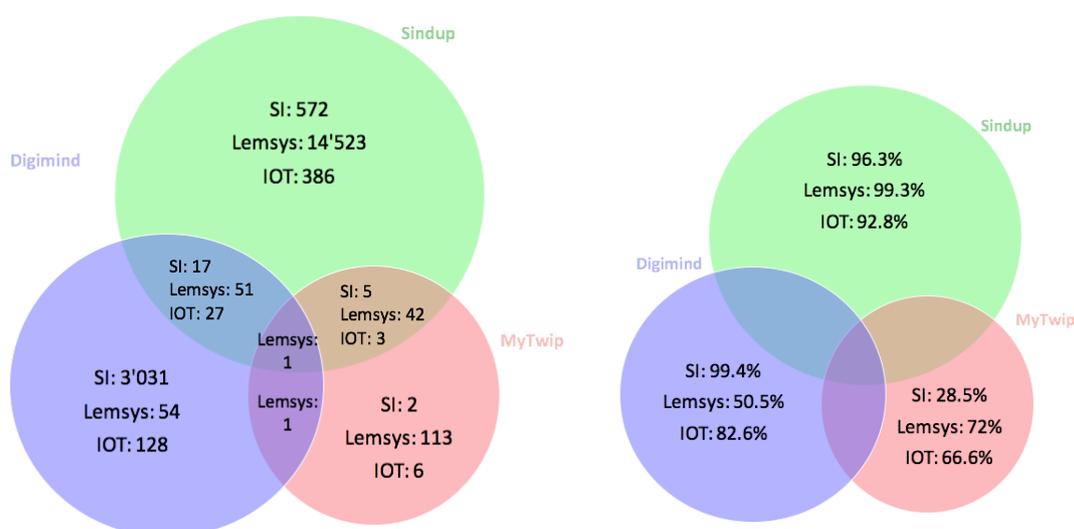
Figure 19 : SA : Complémentarité des résultats retournés



Le premier diagramme de Venn ci-dessus nous montre que sur l'ensemble des résultats collectés via les sources automatiques (SA), Sindup possède 15'481 résultats uniques (qui n'ont été renvoyés ni par Digimind, ni par MyTwip), Digimind en possède 3'213 et MyTwip 121.

Sindup est la plateforme qui partage le plus de résultats en commun avec ses deux concurrentes (95 avec Digimind et 50 avec MyTwip). MyTwip et Digimind quant à elles ne partagent ensemble qu'un résultat. Nous observons également que seul un résultat est commun aux trois plateformes. Ces données nous permettent de calculer le taux de complémentarité des trois plateformes. MyTwip est la plateforme la moins complémentaire avec 70% de résultats qui lui sont uniques. Bien que le taux de complémentarité de Digimind et Sindup soit très proche, c'est cette dernière qui propose le plus de résultats uniques (99% contre 97% pour Digimind). Malgré qu'elle soit la plateforme ayant le plus de données en commun avec les deux autres, ce taux élevé est dû au très grand nombre de résultats récoltés par rapport à ses deux concurrentes.

Figure 20 : SA : Complémentarité des résultats retournés par sujet



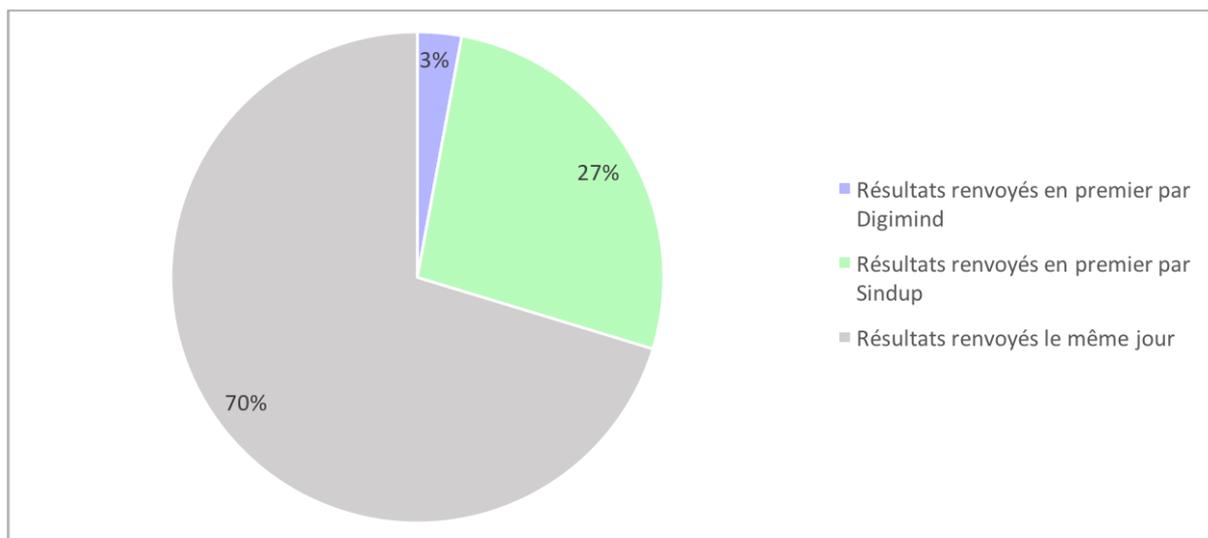
En analysant cet indicateur par sujet de veille, nous observons la même tendance que précédemment : Digimind et Sindup partagent le plus de résultats, ensuite il s'agit de Sindup avec MyTwip, et pour finir Digimind avec MyTwip. Alors que pour les sujets Lemsys (semi-conducteur de puissance) et IoT Sindup possède le plus de résultats uniques (respectivement 99.3% et 92.8%), ce n'est pas le cas pour les Sciences de l'information (SI) où il s'agit de Digimind (99.4% alors que Sindup en a 96.3%). Par ailleurs, MyTwip est la plateforme ayant la plus grande proportion de résultats communs avec ses deux concurrentes pour les sujets de veille SI et IoT (28.5% SI et 66.6% IoT), mais ce n'est pas le cas pour Lemsys (semi-conducteur de puissance) où son taux de complémentarité est plus élevé que celui de Digimind (72% contre 50.5% pour Digimind).

4.5 Rapidité de capture

Cet indicateur qui repose sur la date de collecte des résultats par les plateformes a pour objectif d'identifier si une plateforme récolte un document avant une autre, lorsque le même résultat est retourné par différentes plateformes.

4.5.1 La rapidité de capture des sources manuelles (ou SM)

Figure 21 : SM : Rapidité de capture Digimind VS Sindup

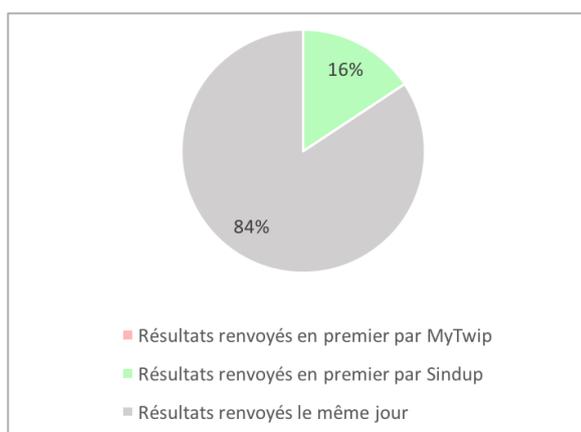


Sur 138 résultats communs¹ renvoyés par Sindup et Digimind, 70% ont été collectés le même jour par les deux plateformes tandis que 27% ont été retournés par Sindup un ou plusieurs jours avant sa concurrente. Seuls 4 résultats représentant 3% ont été collectés en premier par Digimind. Nous notons alors qu'entre Digimind et Sindup, c'est cette dernière qui capture les résultats le plus rapidement, malgré 97 résultats renvoyés la même journée.

Cette même tendance se vérifie pour chacun des sujets de veille. En effet, les résultats retournés le même jour représentent toujours la plus grande proportion et varient entre 65% et 83.3% des résultats retournés. 16.7% à 33.3% des résultats sont collectés par Sindup un ou plusieurs jours avant sa concurrente, et cette dernière ne renvoie que très peu de résultats avant Sindup (entre 0% et 4.4%).

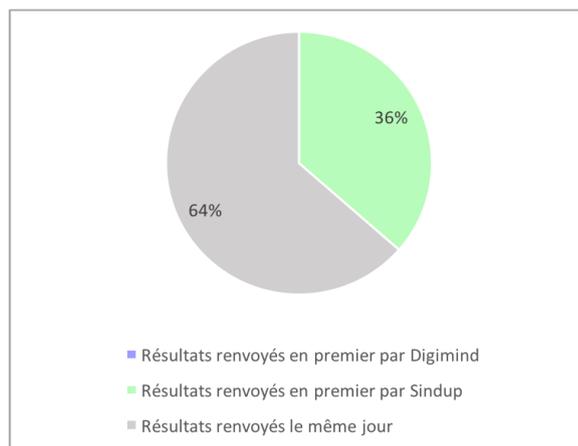
4.5.2 La rapidité de capture des sources automatiques (ou SA)

Figure 22 : SA : Rapidité de capture MyTwip VS Sindup



¹ Ayant nécessité d'être échantillonnée, nous n'avons pas pu comptabiliser les dates de collecte de la requête SI_SM_Axe1_R1 dans ce calcul, Digimind ne permettant pas cette fonctionnalité.

Figure 23 : SA : Rapidité de capture Digimind VS Sindup



Dans la partie précédente présentant la complémentarité des résultats, nous avons pu observer que pour les sources automatiques (SA), Sindup était la plateforme partageant le plus de résultats avec ses deux concurrentes, Digimind et MyTwip. Que ce soit avec l'une ou l'autre des plateformes, c'est la seule à renvoyer des résultats un ou plusieurs jours avant ses concurrentes. En effet, sur 51 résultats communs avec MyTwip, elle en a collecté 16% avant cette dernière (8 résultats). Les 43 autres ont été capturés le même jour. Face à sa concurrente Digimind, sur 96 résultats communs, elle en a renvoyé 36% avant et 64% le même jour. Tout comme pour les SM, en ce qui concerne les SA, Sindup paraît donc être également la plateforme la plus rapide concernant la capture des résultats.

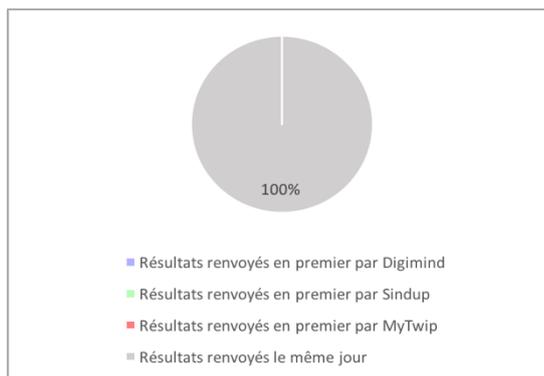
Cette tendance se vérifie également pour chacun des sujets de veille où l'on peut observer que la majorité des résultats a été retournée le même jour par les plateformes. Si une plateforme a été plus rapide qu'une autre, il s'agit toujours de Sindup. Celle-ci a collecté entre 0% et 53% de résultats (selon le sujet de veille et la plateforme comparée) un ou plusieurs jours avant l'une de ses concurrentes Digimind ou MyTwip.

Figure 24 : SA : Rapidité de capture Digimind VS MyTwip



Digimind et MyTwip possède un seul résultat commun uniquement à elles deux. Ce résultat a été retourné le même jour par les plateformes.

Figure 25 : SA : Rapidité de capture Digimind VS MyTwip VS Sindup



Il en est de même pour l'unique résultat communs aux trois plateformes, qui lui aussi, a été retourné à la même date.

4.6 Tableaux récapitulatifs des performances des plateformes

Ces tableaux permettent de visualiser les performances des différentes plateformes, tout d'abord selon l'approche utilisant les sources manuelles (SM), puis selon les sources automatiques (SA).

Pour chaque indicateur, la plateforme la plus performante est celle qui obtient le plus grand nombre de points. Dans notre calcul, nous estimons que tous les indicateurs ont la même importance, ce qui, dans la réalité, dépend de la perception et des besoins de chacun.

4.6.1 Récapitulatif pour les sources manuelles (ou SM)

Pour chaque indicateur, la plateforme la plus performante obtient 1 point. Ainsi, la plateforme obtenant le score le plus élevé est celle qui obtient les meilleurs résultats.

Tableau 10 : Tableau récapitulatif pour les sources manuelles

| | Digimind | Sindup |
|---------------------|----------|--------|
| Volumétrie | 1 | 0 |
| Pertinence | 1 | 0 |
| Précision | 1 | 0 |
| Complémentarité | 1 | 0 |
| Rapidité de capture | 0 | 1 |
| TOTAL | 4 | 1 |

Lorsque l'on contraint les plateformes à renvoyer des résultats selon des sources sélectionnées et insérées manuellement, la plateforme Digimind se révèle être meilleure que sa concurrente Sindup, qui est en revanche plus performante lorsqu'il s'agit de la rapidité de capture.

4.6.2 Récapitulatif pour les sources automatiques (ou SA)

Pour chaque indicateur, la plateforme la plus performante obtient 2 points et la deuxième plus performante, 1 point. Ainsi, la plateforme obtenant le score le plus élevé est celle qui obtient les meilleurs résultats.

Tableau 11 : Tableau récapitulatif pour les sources automatiques

| | Digimind | Sindup | MyTwip |
|---------------------|----------|--------|--------|
| Volumétrie | 1 | 2 | 0 |
| Pertinence | 1 | 2 | 0 |
| Précision | 0 | 1 | 2 |
| Complémentarité | 1 | 2 | 0 |
| Rapidité de capture | 1 | 2 | 0 |
| TOTAL | 4 | 9 | 2 |

Lorsque les plateformes renvoient des résultats sans indication pour le choix des sources, mais au contraire en utilisant leur propre base de données, alors Sindup se trouve être plus performante que ses concurrentes. Cependant, MyTwip se révèle être la plateforme permettant d'obtenir la meilleure précision.

5. Conclusion

5.1 Réponses à nos questions de recherche

Pour rappel, nos deux questions de recherche étaient les suivantes :

- Quelle méthode et quels indicateurs permettent d'évaluer la qualité des plateformes de veille Digimind, Sindup et MyTwip sur la base des résultats retournés ?
- Quelle est la plateforme la plus performante au regard de cette évaluation ?

5.1.1 La méthode d'évaluation employée

Dans notre étude, nous avons proposé une méthode d'évaluation mixte alliant des données quantitatives à des données qualitatives. Trois sujets de veille très différents ont été définis de sorte que l'on puisse comparer les plateformes et généraliser les résultats à l'ensemble des thématiques existantes. Par ailleurs, nous avons essayé de paramétrer les plateformes de la même manière, de leur intégrer les mêmes requêtes et d'analyser les résultats retournés par chacune d'entre elles.

Différents indicateurs quantitatifs et qualitatifs ont été définis pour évaluer la qualité des résultats des plateformes : la volumétrie, la pertinence, la précision, la complémentarité et la rapidité de capture.

5.1.2 La performance des plateformes

Comme décrit au point 4.6 Tableaux récapitulatifs des performances des plateformes, nous observons que la plateforme de veille la plus performante n'est pas la même lors de l'interrogation des sources automatiques et des sources manuelles.

En effet, d'après le système de ranking que nous avons mis en place, Digimind semble plus performante que sa concurrente Sindup lorsque les plateformes interrogent des sources que nous avons sélectionnées (SM). Elle est meilleure selon l'ensemble des indicateurs définis sauf en ce qui concerne la rapidité de capture où Sindup la devance.

Lorsque les plateformes n'ont pas de sources qui leur sont imposées (SA), c'est en revanche Sindup qui semble largement plus performante que ses concurrentes MyTwip et Digimind. Le seul indicateur pour lequel elle ne figure pas au premier rang est la précision pour laquelle elle est devancée par MyTwip.

5.2 Limites de la méthode et des résultats obtenus

5.2.1 Le paramétrage des plateformes

Comme décrit plus en détails dans le chapitre 3.4 *Paramétrage des plateformes*, le paramétrage des trois plateformes est très différent d'un outil à l'autre, ce qui est particulièrement le cas de la plateforme MyTwip. Ainsi, bien que nous ayons essayé de les paramétrer de la manière la plus identique possible, nous ne pouvons pas affirmer qu'elles sont en tout point comparables.

Par ailleurs, et pour permettre leur comparaison, nous n'avons pas utilisé l'ensemble des fonctionnalités proposées mais uniquement les options communes aux trois plateformes. Notre évaluation n'est donc pas basée sur le potentiel complet des différents outils et sur les options à disposition qui représentent parfois leur valeur ajoutée.

5.2.2 Le choix des indicateurs

Pour définir la qualité des plateformes, d'autres indicateurs auraient pu être analysés, tels que :

- l'analyse des doublons (la période de collecte ayant mis en lumière le fait que certaines plateformes semblent rapporter plus de doublons que d'autres),
- la fraîcheur des informations (c'est-à-dire la date de publication des informations sur le web, certaines plateformes renvoyant des résultats publiés depuis plusieurs mois et de fait certainement moins pertinents dans le cadre d'une véritable veille).

Cependant, au vu du temps à disposition et pour des raisons de faisabilité, nous n'avons pas pu les développer dans le cadre de notre étude.

Egalement, l'indicateur "rapidité de capture" aurait pu être affiné si nous avions relevé, en plus de la date, l'heure de collecte des résultats. En effet, nous nous sommes aperçues durant l'analyse que le relevé de la date ne permettait pas de distinguer suffisamment d'écart entre les trois plateformes.

Par ailleurs, l'analyse de la pertinence est certainement l'indicateur induisant le biais le plus important, pour plusieurs raisons :

- La pertinence est en soit une notion subjective, un document pertinent pour une personne ne le sera pas pour une autre.
- Pour des raisons de faisabilité, nous n'avons pas pu le faire à l'aveugle (de manière à ignorer de quelle plateforme proviennent les résultats), ce qui peut biaiser nos analyses car il existe potentiellement des préférences pour une plateforme plutôt qu'une autre.
- Enfin, le fait de ne pas correctement maîtriser les domaines couverts par les sujets de veille IoT et Lemsys (semi-conducteur de puissance) a également une incidence directe sur notre évaluation de la pertinence des résultats collectés.

5.2.3 Autres éléments de notre méthode

De plus, la méthode que nous proposons a exigé de nombreuses manipulations de notre part, que ce soit lors de l'intégration des requêtes dans les plateformes, lors de la collecte des résultats ou de l'analyse de ceux-ci, ce qui a une incidence sur le risque d'erreur.

La période de la collecte peut être également discutée. En effet, pour des raisons de faisabilité nous avons dû la mettre en place dès le mois d'août, or nous savions qu'il s'agit d'un mois moins riche en actualités.

5.2.4 Les résultats

Les quantités variables de résultats obtenus via les plateformes ont posé quelques difficultés à l'analyse :

- Avec trop peu de résultats, comme cela a par exemple été le cas pour l'approche "sources manuelles" (SM) du sujet sur l'IoT, il est difficile de dégager des conclusions sur la qualité des plateformes,
- Avec un trop grand nombre de résultats et en particulier une grande disparité entre les plateformes comme pour l'approche "sources automatiques" (SA) de Lemsys (semi-conducteur de puissance), il est également délicat de mener une comparaison. En effet, s'il est d'une part impossible d'analyser l'ensemble des résultats retournés, d'autre part, les échantillons ne permettent pas de réaliser l'ensemble des analyses souhaitées (en particulier concernant uniquement les résultats pertinents).

Par ailleurs, les résultats liés aux indicateurs que nous avons définis sont à manipuler avec précaution. Par exemple, la précision ne prend pas en compte le bruit documentaire généré et la complémentarité comprend les résultats pertinents et non pertinents.

De plus, lors de notre étude et de l'attribution des rangs aux trois plateformes, nous avons estimé que chaque indicateur possède la même importance et nous leur avons donc attribué la même pondération. Or, dans la réalité, cela dépend de la perception et des besoins du veilleur.

Finalement, notre évaluation et nos résultats, bien que nous ayons observé certaines tendances, reposent uniquement sur trois sujets de veille. Les résultats auraient peut-être été différents en choisissant d'autres thématiques.

5.3 Proposition d'autres méthodes d'évaluation

Afin de supprimer le biais lors de l'analyse des résultats dû à la préférence pour une plateforme plutôt qu'une autre, il serait intéressant de reprendre notre méthodologie en procédant à une analyse à l'aveugle. Nécessitant plus de temps, cette nouvelle méthode pourrait se dérouler selon les étapes ci-dessous :

- Collecte des résultats par un veilleur dans différents tableurs renseignant les plateformes à l'origine des données collectées,
- Report de ces données dans un seul tableur par requête (sans indication quant à la plateforme d'origine),
- Analyse des résultats par un second veilleur,
- Croisement des données par le premier veilleur, avec renseignement des plateformes d'origine.

Une autre limite évoquée à propos de notre méthodologie concerne les options des plateformes, que nous n'avons pas prises en compte dans un souci de comparaison. Celles-ci permettent pourtant de prendre la mesure de l'entière capacité des plateformes. Une méthodologie possible serait alors de répéter notre étude en optimisant le paramétrage des trois plateformes (utilisation de toutes les options utiles). De cette manière, il serait en outre

ainsi possible de comparer MyTwip aux deux autres plateformes concernant notre approche via les “sources manuelles”, l’équipe technique de Coexel étant en mesure de paramétrer au mieux leur outil.

Par ailleurs, l’approche par les “sources manuelles” nous a permis de relever que les plateformes Digimind et Sindup ne collectent pas les mêmes résultats avec pourtant les mêmes sources à disposition. Si nous n’avons pas accès à leurs algorithmes, il serait cependant intéressant d’analyser les résultats retournés de manière à en déduire certains fonctionnements et comprendre ces différences. Il serait par exemple envisageable d’analyser :

- les sites utilisés par les plateformes (communs et différents),
- les URL renvoyées (communes et différentes),
- les mots-clés présents ou absents des résultats,
- etc.

6. Bibliographie

ASSELIN, Christophe, 2015. La veille stratégique musclée par les médias sociaux : Digimind transforme la Competitive intelligence. *Digimind Blog* [en ligne]. 2 décembre 2015. [Consulté le 27 avril 2017]. Disponible à l'adresse : <https://blog.digimind.com/fr/fr-int/digimind-intelligence-nouvelle-interface-nouvelles-fonctions-sociales/>

ASSOCIATION DES PROFESSIONNELS DE L'INFORMATION ET DE LA DOCUMENTATION, 2015. Veille. *ADBS: l'association des professionnels de l'information et de la documentation* [en ligne]. [Consulté le 22 mars 2017]. Disponible à l'adresse : http://www.adbs.fr/veille-19022.htm?RH=OUTILS_VOC

ASSOCIATION FRANÇAISE DE NORMALISATION, 1998. *Prestations de veille et prestations de mise en place d'un système de veille* [en ligne]. 1998. S.I. : AFNOR. [Consulté le 22 mars 2017]. Disponible à l'adresse : http://quoniam.info/competitive-intelligence/PDF/ebooks/Norme_Francaise_Prestations_de_Veille.pdf

AUDEH, Bissan, BEAUNE, Philippe et BEIGBEDER, Michel, 2015. MOR : Mesure orientée rappel pour les systèmes de recherche d'information. *Document numérique*. 23 avril 2015. Vol. 18, n° 1, p. 37-54.

BACCINI, Alain, DESJEAN, Sébastien, KOMPAORÉ, Désiré et MOTHE, Josiane, 2010. Analyse des critères d'évaluation des systèmes de recherche d'information. *Revue des Sciences et Technologies de l'Information*. 2010. Vol. 29, n° 3, p. 289-308. Accès https://www.irit.fr/publis/SIG/2010_ISI_BADSKNDMJ.pdf

BACHR, Ahmed, 2008. La veille documentaire : Session 3 : Réaliser une veille manuelle. *Slideshare.net* [en ligne]. 12 mars 2008. [Consulté le 29 décembre 2017]. Disponible à l'adresse : https://fr.slideshare.net/Bachr/veille-documentaire?next_slideshow=1

BALMISSE, Gilles et MEINGAN, Denis, 2008. *La veille 2.0 et ses outils*. [S.I.] : Hermes Science publ. ; Paris : Lavoisier. ISBN 978-2-7462-1929-8

BEZENÇON, Christophe et ROULET, Aurélie, 2014. *Evaluation de la performance et de l'impact de la veille* [en ligne]. Genève : Haute école de gestion de Genève. Mémoire de recherche. [Consulté le 9 mars 2017]. Disponible à l'adresse : <http://doc.rero.ch/record/209591>

BOISARD, Vincent, 2016. *Mytwip : solution professionnelle de veille* [document PDF]. 2016. Support de cours : Cours "784-2n Veille stratégique 2", Haute école de gestion de Genève, filière Information Documentaire, année académique 2017-2018.

BRIAND, Bénédicte, BUFFETEAU-HEJBLUM, Annie, CUDELOU, Jean-François, DROUET, Claire, MANGEOT, Annie, PAILLARD, Marie-Danièle et REGNAULT, Madeleine, 2009. Indicateurs de performance des services documentaires. *Documentaliste-Sciences de l'Information*. 1 février 2009. Vol. 39, n° 1, pp. 26-33.

CACALY, Serge et LE COADIC, Yves-François, 2004. *Dictionnaire de l'information*. 2e éd. Paris : A. Colin. ISBN 978-2-200-26682-0

CLUB WEBTECHNO AEGE, 2017. Comparatif de logiciels de veille. *Portail de l'IE* [en ligne]. [Consulté le 8 juillet 2017]. Disponible à l'adresse : <https://portail-ie.fr/resource/autre/1531/comparatif-de-logiciels-de-veille>

COEXEL, 2016. *Mytwip: la solution globale pour optimiser vos activités de veille technologique et stratégique* [en ligne]. [Consulté le 18 mars 2017]. Disponible à l'adresse : <http://www.coexel.com/>

COEXEL, [2017a]. A propos. *Coexel* [en ligne]. [Consulté le 27 juin 2017]. Disponible à l'adresse : <http://www.coexel.com/a-propos/>

COEXEL, [2017b]. Offres. *Coexel* [en ligne]. [Consulté le 27 juin 2017]. Disponible à l'adresse : <http://www.coexel.com/offres/>

- COEXEL, [2017c]. Services. *Coexel* [en ligne]. [Consulté le 27 juin 2017]. Disponible à l'adresse : <http://www.coexel.com/services/>
- COHEN, Corine, 2004. *Veille et intelligence stratégiques*. Paris : Hermès Science Publications. Management, finance et gestion. ISBN 978-2-7462-0851-3
- COMMISSARIAT GÉNÉRAL DU PLAN, 1994. Rapport du groupe: « Intelligence économique et stratégie des entreprises ». *La documentation française* [en ligne]. [Consulté le 23 mars 2017]. Disponible à l'adresse : <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/074000410.pdf>
- DEBONNE, Eric, 2012. Solution dans la course à l'innovation. *Archimag.com. Guide pratique*. Décembre 2012. No 47, pp.56-57. 2260-1708
- DELENGAIGNE, Xavier, 2012. *Organiser sa veille sur internet: Au-delà de Google...Outils et astuces pour le professionnel* [en ligne]. Paris : Eyrolles [Consulté le 20 mai 2017]. Disponible à l'adresse : <http://hesge.scholarvox.com/catalog/book/docid/88818042?searchterm=organiser%20sa%20veille%20sur%20internet> [accès par abonnement]
- DESFRICHE DORIA, Orélie, 2015. Le rôle de l'activité dans l'évaluation de l'information, The role of activity in information evaluation. *Document numérique*. 23 avril 2015. Vol. 18, n° 1, p. 9-36.
- DIGIMIND, 2011. Solutions de veille stratégique : guide de benchmark 2011 [document PDF]. *Digimind* [en ligne]. [Consulté le 08 juillet 2017]. Disponible à l'adresse : <http://www.digimind.com/fr/resources/fiches-produit/logiciels-de-veille-200-criteres-pour-mieux-choisir/>
- DOUCET, Christian, 2010. *La qualité*. 3e éd. mise à jour. Paris : Presses universitaires de France. Que sais-je?, 2779. ISBN 978-2-13-057896-3
- DUALÉ, Robin, 2010. *Management stratégique de l'information*. Paris : Weka. ISBN 978-2-7337-0444-8
- DUPIN, Corinne, 2014. *Guide pratique de la veille*. Mont Saint-Aignan : Klog. ISBN 979-10-92272-00-0
- Editeurs de solutions de veille et e-réputation [document PDF]. *Archimag* [en ligne]. Novembre 2015. [Consulté le 08 juillet 2017]. Disponible à l'adresse : <http://www.archimag.com/sites/archimag.com/files/Archimag289-Logiciel-Veille.pdf>
- Evaluer. Larousse [en ligne]. [sans date]. [Consulté le 7 août 2017]. Disponible à l'adresse : <http://www.larousse.fr/dictionnaires/francais/%C3%A9valuer/31795>
- FERRIERE, Pierre, 2015. Surveiller ses réseaux sociaux avec la nouvelle solution de veille Digimind. *Archimag* [en ligne]. [Consulté le 27 avril 2017]. Disponible à l'adresse : <http://www.archimag.com/veille-documentation/2015/03/24/surveiller-reseaux-sociaux-nouvelle-solution-veille-digimind>
- FOENIX-RIOU, Béatrice, 2014. Pistes et outils pour étoffer son sourcing. *Recherche éveillée* [en ligne]. 13 avril 2014. [Consulté le 1 septembre 2017]. Disponible à l'adresse : <http://www.recherche-eeveillee.com/2014/04/pistes-et-outils-pour-etoffer-son-sourcing.html>
- FOSSO, Arnaud Noël, 2012. Veille et Intelligence économique: quelles différences? *Repère de l'information-documentation* [en ligne]. 2012. [Consulté le 23 mars 2017]. Disponible à l'adresse : <http://fossoarnaud.unblog.fr/2012/12/16/veille-et-intelligence-economique-quelles-differences/>
- FROCHOT, Didier, 2006. Comment mettre en place un dispositif de veille. *les-ifostrateges.com* [en ligne]. [Consulté le 11 août 2017]. Disponible à l'adresse : <http://www.les-ifostrateges.com/article/0606267/comment-mettre-en-place-un-dispositif-de-veille>

- GOBEILL, Julien, [2016]. *Evaluer la recherche d'information (RI)* [Document PDF].
- Support de cours : cours "Fondements des sciences de l'information", Haute école de gestion de Genève, filière information documentaire, année académique 2016-2017
- HASSEN, Bekkadour, 2014. Veille strategique et intelligence economique. *LinkedIn Slideshare* [en ligne]. [Consulté le 23 mars 2017]. Disponible à l'adresse : <https://fr.slideshare.net/bekaddour/veille-strategique-et-intelligence-economique-cas-dentreprise-algerie-telecom>
- INNOVARC, [2016]. Plateforme de veille webSO+. In : *InnovARC* [en ligne]. [Consulté le 18 mars 2017]. Disponible à l'adresse : <http://www.innovarc.eu/innovarc/consortiums-projets-et-success-stories/>
- KHAKIMOV, Ildar. 2013. Google Alerts VS Mention VS Talkwalker. *Moz* [en ligne]. [Consulté le 1 juillet 2017]. Disponible à l'adresse : <https://moz.com/ugc/google-alerts-vs-mention-vs-talkwalker>
- LAFOUGE, Thierry, LE COADIC, Yves-François et MICHEL, Christine, 2002. *Eléments de statistique et de mathématique de l'information: infométrie, bibliométrie, médiométrie, scientométrie, muséométrie, webométrie: cours avec exemples et exercices corrigés*. Villeurbanne : Presses de l'ENSSIB. Les cahiers de l'ENSSIB 001. ISBN 978-2-910227-37-1
- LESCA, Nicolas et CARON FASAN, Marie-Laurence, 2006. *Veille anticipative: une autre approche de l'intelligence économique*. Paris : Hermès Science. Management et informatique. ISBN 978-2-7462-1459-0
- L'OEIL AU CARRE, 2013. Cartographie des principaux types de veille. *MindMeister* [en ligne]. [Consulté le 20 août 2017]. Disponible à l'adresse : <https://www.mindmeister.com/fr/56791764?title=cartographie-des-principaux-types-de-veille>
- MADINIER, Hélène, 2016. *Le cycle du renseignement : cours théorique 2* [document PDF]. Année académique 2016-2017.
- Support de cours : Cours "723-13nf Veille stratégique 1", Haute école de gestion de Genève, filière Information Documentaire, année académique 2016-2017
- MADINIER, Hélène, 2017. *Veille 2: cours 1* [document PDF].
- Support de cours : cours "Veille stratégique 2", Haute école de gestion de Genève, filière Information documentaire, année académique 2016-2017
- MADINIER, Hélène et GAUDINAT, Arnaud, 2017. *Sujets de veille pour les 2 travaux de recherche Master sur l'évaluation des sources de veille/des outils de veille* [fichier Microsoft Word]. 14 mars 2017.
- Document interne à la Haute école de gestion, Genève
- MANNING, Christopher D., RAGHAVAN, Prabhakar et SCHÜTZE, Hinrich, 2008. Evaluation in information retrieval. *Introduction to Information Retrieval* [en ligne]. Cambridge : Cambridge University Press. Accès <https://nlp.stanford.edu/IR-book/pdf/08eval.pdf>
- MANISCALCO, Elisa, 2016. *Etude de la mise en place d'un service de veille pour le corps enseignant de la HEdS* [en ligne]. Genève : Haute école de gestion de Genève. Travail de Bachelor. [Consulté le 22 mars 2017]. Disponible à l'adresse : <http://doc.rero.ch/record/278090>
- NIE, Jian-Yun, 2016a. Introduction à la RI. *Université de Montréal* [en ligne]. 2016. [Consulté le 12 mai 2017]. Disponible à l'adresse : <http://www.iro.umontreal.ca/~nie/IFT6255/Introduction.pdf>
- NIE, Jian-Yun, 2016b. Le domaine de recherche d'information: un survol d'une longue histoire. *Université de Montréal* [en ligne]. 2016. [Consulté le 13 mai 2017]. Disponible à l'adresse : <http://www.iro.umontreal.ca/~nie/IFT6255/historique-RI.pdf>

Panorama de solutions de veille [document PDF]. *Archimag* [en ligne]. Juin 2017. [Consulté le 08 juillet 2017]. Disponible à l'adresse : http://www.archimag.com/sites/archimag.com/files/adds/Archimag-305_veille_collab.pdf

PINEL-SAUVAGNAT, Karen et MOTHE, Josiane, 2013. Mesures de la qualité des systèmes de recherche d'information. *Ingénierie des systèmes d'information* [en ligne]. 28 juin 2013. Vol. 18, n° 3, p. 11-38. DOI 10.3166/isi.18.3.11-38. Accès <http://isi.revuesonline.com/article.jsp?articleId=18650>

REBMANN, Odile, 2016. Digimind leader des outils de veille en matière de satisfaction client. *Blueboat* [en ligne]. 18 octobre 2016. [Consulté le 27 avril 2017]. Disponible à l'adresse : <http://www.blueboat.fr/digimind-leader-outils-veille-satisfaction-client/>

RUCH, Patrick, [2016]. *Échantillonnage* [présentation powerpoint]. Année académique 2016-2017.

Support de cours : Cours "M4C2 - Statistiques appliquées à la recherche", Haute école de gestion de Genève, filière Information Documentaire, année académique 2016-2017.

SALA, Jacqueline, 2016. Digimind intègre la reconnaissance d'images à sa veille sociale. *Veillemag.com* [en ligne]. [Consulté le 27 mars 2017]. Disponible à l'adresse : http://www.veillemag.com/Digimind-integre-la-reconnaissance-d-images-a-sa-veille-sociale_a3041.html.

SCHWOB, Elodie, 2017a. *Digimind : une plateforme de veille professionnelle* [document PDF]. Printemps 2017.

Support de cours : Cours "784-2n Veille stratégique 2", Haute école de gestion de Genève, filière Information Documentaire, année académique 2016-2017.

SCHWOB, Elodie, 2017b. *Mytwip : solution de veille professionnelle* [document PDF]. 2016-2017.

Support de cours : Cours "784-2n Veille stratégique 2", Haute école de gestion de Genève, filière Information Documentaire, année académique 2016-2017.

SCHWOB, Elodie, [sans date]. *Sindup : plateforme de veille stratégique et e-réputation* [document PDF]. Printemps 2017.

Support de cours : Cours "784-2n Veille stratégique 2", Haute école de gestion de Genève, filière Information Documentaire, année académique 2016-2017.

Sindup, l'étoile montante de la veille. *GFII* [en ligne]. [Sans date]. [Consulté le 25 mars 2017]. Disponible à l'adresse : <http://www.gfii.fr/fr/document/Sindup-l-etoile-montante-de-la-veille>

SINDUP, [sans date a]. La société - Sindup. *Sindup* [en ligne]. [Consulté le 27 juin 2017]. Disponible à l'adresse : <http://fr.Sindup.com/societe>

SINDUP, [sans date b]. Offres. *Sindup* [en ligne]. [Consulté le 27 juin 2017]. Disponible à l'adresse : <http://fr.Sindup.com/offres>

SINDUP, 2017. *Sindup* [en ligne]. 2017. [Consulté le 18 mars 2017]. Disponible à l'adresse : <http://fr.Sindup.com/>

Solutions dans la course à l'innovation. *Archimag guide pratique*. Décembre 2012. No 47, pp.56-79. ISSN 2260-1708

Source (information). *Techno-science.net* [en ligne]. [Sans date]. [Consulté le 7 août 2017]. Disponible à l'adresse : <http://www.techno-science.net/?onglet=glossaire&definition=10708>

SUTTER, Eric, 2006. *L'évaluation et les indicateurs de la performance des activités info-documentaires*. Paris : ADBS. L'essentiel sur... ADBS. ISBN 978-2-84365-082-6.

TISSERAND-BARTHOLE, Carole, 2015. Gratuit contre payant : la performance des plateformes de veille sur les médias sociaux - une enquête à suivre [document PDF]. *Netsources* [en ligne]. Juillet/Août 2015. [Consulté le 27 mars 2017]. Disponible à l'adresse : <https://www.fla-consultants.com/images/NetSources117.pdf>

THOMAS, Armelle, 2008. Parce que la veille bouge. *Documentaliste-Sciences de l'Information*. 28 décembre 2008. Vol. 45, n° 4, p. 30–31.

UNIVERSITÉ EUROPÉENNE DE BRETAGNE, 2016. Veille professionnelle et outils collaboratifs. *Cultures numériques dans l'enseignement* [en ligne]. [Consulté le 22 mars 2017]. Disponible à l'adresse : <http://www.ressources.univ-rennes2.fr/cultures-numeriques-dans-l-enseignement/veille/1-quest-ce-que-la-veille/1-1-definitions-de-la-veille/>

UNIVERSITÉ DE FRANCHE-COMTÉ (UFC) [et al.], 2017. Résumé du projet INTERREG Plateforme de veille multifonctionnelle webSO+ [fichier Microsoft Word]. 20 janvier 2017.

Document interne à la Haute école de gestion, Genève

WINELLO, 2011. Glossaire des termes utilisés dans la veille. *En veille* [en ligne]. [Consulté le 26 mars 2017]. Disponible à l'adresse : <http://www.enveille.com/index.php/glossaire/Glossaire-sur-la-veille-2/P/Plateforme-de-veille-9/>

WOODWARD, Matthew, 2013. Web Monitoring Software Wars - TalkWalker Alerts vs Google Alerts. *Matthew Woodward* [en ligne]. 13 mars 2017. [Consulté le 1 juillet 2017]. Disponible à l'adresse : <http://www.matthewwoodward.co.uk/experiments/which-is-the-best-web-monitoring-tool-talkwalker-vs-google-alerts/>