
Modélisations dans une approche de veille générique (GWatch) :

Clustering centré acteurs de veille

Mseddi Rim¹, Sahbi Sidhom², Malek Ghenima³ et Henda Ben
ghezela⁴

1. RIADI-GDL-Manouba, Email : b_mseddirim@yahoo.fr
2. LORIA-Université de lorraine, France, Email : sahbi.sidhom@loria.fr
3. ESCEM-Manouba, Email : malek.ghenima@esem.rnu.tn
4. RIADI-GDL-Manouba, Email: henda.hg@cck.rnu.tn

RESUME. Dans le cadre de notre recherche sur la modélisation de l'acteur de veille et le développement d'un système de veille générique centré sur l'acteur, nous présentons dans cet article une nouvelle approche d'intégration semi-automatique d'un processus de veille associé à un système de recherche d'information existant. Cette approche vise essentiellement à faciliter la mise en place d'un système de veille, ce qui permettra, d'un côté, la définition des besoins de veille à partir des informations collectées et, d'un autre côté, la structuration puis l'évolution des projets de veille dans une approche systémique avec l'information traitée.

ABSTRACT. As part of our research on watch actor modeling and the development of a generic watch system centered on the actor, we present in this paper a new semi-automatic integration approach to watch process with existing information retrieval system. This approach is essentially to facilitate the establishment of a watch system associated, which will, on the one hand, the definition of the watch needs from information collected and, on the other hand, the structure and the evolution of intelligence projects in a systemic approach with processed information.

MOTS-CLES : système de veille, recherche d'information, gestion de la connaissance, intégration, modélisation de l'acteur, clustering.

KEYWORDS: Watch system, information research, Knowledge organisation, integration, actor modeling, clustering.

1. Introduction

Si la veille a maintenant acquis ses lettres de noblesse, elle reste, comme tout concept récent, encore mal comprise » affirme J.P. Bernat (Bernat, 2008). Il existe plusieurs définitions de la veille qui peuvent concerner soit les ressources, soit les outils ou les types de veilles utilisés. Le plus souvent, elle est considérée comme une activité, démarche ou ensemble de méthodes et de techniques de gestion de l'information à valeur ajoutée (définition des besoins, recherche, collecte, traitement et diffusion). On distingue aussi entre la veille qui utilise des ressources et des outils hypertextuels, web natif, web 2.0 ou web conceptuel (Delaby, 2004; Leitzelman, 2009a; Pateryon, 1998; Raquel, 2011).

Dans un processus de veille, la première étape à savoir la définition des besoins de veille reste la tâche la plus difficile pour un acteur de veille d'autant lorsque ces besoins sont imprécis ou lorsque le corpus ne lui est pas familier. Dans notre contexte, l'objectif est de développer le système de veille afin de répondre le plus pertinemment possible. Pour la veille, nous considérons trois types d'information comme ressources au processus: le jugement de la pertinence par l'acteur de veille, des clusters récupérés à partir du corpus initial retournés par le système de recherche d'information et des données globales introduites explicitement par les différents acteurs.

Le premier type repose sur le retour de pertinence explicite de l'acteur de veille à travers la sélection des documents retenus dans des projets de veilles antérieurs. Le deuxième type repose sur l'analyse du corpus afin d'établir des relations entre les clusters et les environnements de veille de l'acteur. Pour finir, le dernier type encapsule toutes les informations explicites qui alimentent le système de veille. Elles peuvent parvenir de l'animateur, du veilleur ou même de l'utilisateur final à travers la sélection de ces centres d'intérêt, des environnements de veille, etc.

Dans cet article, nous présentons, dans le premier paragraphe, la gestion des informations dans un processus de veille. Dans un deuxième paragraphe, nous décrivons notre approche et nous détaillons, dans le dernier paragraphe, nos résultats d'analyse. Nous finirons par une conclusion et la présentation de quelques perspectives.

2. La gestion des informations dans un processus de veille

Les outils de veille intégrés proposés sur le marché diffèrent selon les ressources gérées et les outils de traitement de l'information proposés. En général, ils proposent la même démarche qui débute par la définition des besoins sous forme de requêtes qui peuvent être soit ponctuelle (projet), cyclique (rapport) ou continue (alerte). Ces requêtes peuvent concerner des sources spécifiques ou un scannage général du web. Les informations retournées sont sous forme de flux RSS, de pages web, de Newsletters.... Certains outils de veille proposent aussi des outils de traduction ou de génération de résumé automatique.

Les informations collectées seront, ensuite, traitées par le veilleur qui va sélectionner l'information pertinente, nettoyer les résultats des données superflues, générer l'information à valeur ajoutée et la partager en diffusion.

Dans la littérature, plusieurs processus de veille ont été proposés, nous pouvons citer les travaux de Lesca et al. qui insiste sur l'importance de la création collective de sens et la mémorisation des connaissances tacites (LESCANNING, VAS-IC) (Lesca, 2011). D'un autre côté, la méthode EQUA²te proposée par Thierry O. et Amos D. (Thiery, 2002) montre l'apport de l'annotation sur des solutions antérieures afin de les associer aux problèmes de recherche d'information actuels. Dans le même contexte, Sidhom propose le modèle SIMBAD (Sidhom, 2002) qui montre, en amont et en aval d'un cycle de veille, l'apport de l'annotation, d'une part, dans la clarification des objectifs en formalisant les indicateurs de recherche d'informations, et d'autre part, la valorisation des informations à valeur ajoutée dans la prise de décision ou l'identification des signaux faibles.

L'aspect collaboratif du processus de veille est très important car à travers des interactions dynamiques qu'on peut assurer une pratique évolutive et constructive des connaissances collectives (Fayard, 2002). Dans cette réflexion, plusieurs méthodes ont été mises en oeuvre afin de faciliter cette dynamique de groupe. Nous pouvons citer la méthode PUZZLE (Raquel, 2011) qui consiste à faire agir les différents acteurs en leur proposant des brèves d'informations. Jakobiak propose aussi une méthode collective pour l'innovation à travers sa théorie des ingrédients (Jakobiak, 2005).

D'autres méthodes se sont focalisées sur la recherche d'information anticipative par l'identification (ou la détection) des signes d'alertes précoces. Ces informations peuvent indiquer un changement possible dans l'environnement surveillé. Dans cette approche, la veille est observée stratégiquement et alimentera une base de connaissances actionnables pour la gestion des risques et des opportunités de l'organisation (Lesca, 2001; Sidhom, 2011). Quant aux retours d'expérience, la pratique de veille dans les organisations et les entreprises montrent des défaillances majeures dans le bon fonctionnement de ce processus, à savoir: le manque d'implication du personnel, la surcharge cognitive, la perte dans l'hyperespace, le manque de pertinence des résultats retournés, etc.

Dans ce qui suit, nous allons développer trois types essentiels d'information qui sont gérés par un système de veille, à savoir : le profil veilleur, le profil document et le profil requête.

2.1. Le profil veilleur

Les études sur le profil veilleur que nous avons examinées dans la littérature grise sont peu nombreuses ce qui peut être expliqué par une focalisation sur les pratiques et les finalités de veille. On oublie que sur cet acteur, repose l'intégralité du processus de veille. Pour définir un veilleur, Kislin rapproche le veilleur à l'analyste en intelligence économique. Il le définit comme « *un décideur particulier qui interagit à la fois dans le domaine de l'information et avec son homologue dans le domaine économique* » (Kislin, 2007). D'un autre côté, Thomas le présente comme

un «*travailleur du savoir qui est capable de surveiller ce qui l'intéresse quotidiennement*» (Thomas, 2008).

Si nous tentons de comparer le profil d'un veilleur et le profil d'un utilisateur d'un système de recherche d'information (SRI), d'une manière générale, nous pourrions déduire qu'un veilleur partage quelques caractéristiques avec cet utilisateur (informations générale ou spécifique, centres d'intérêt, communautés,...), mais présente des spécificités qui dépendent directement de :

- Son rôle dans le processus de veille : un veilleur peut être un animateur, veilleur ou utilisateur final d'un projet de veille. Ce rôle peut changer d'un projet à un autre. Un veilleur va, donc, avoir des droits d'accès, de consultation (à des dossiers sensibles) et de traitement de l'information qui diffèrent d'un projet à un autre.

- La sélection des informations pertinentes : dans un SRI des méthodes ont été établies afin de calculer la pertinence d'une source par rapport à un utilisateur comme le calcul de nombre de clique, le temps passé sur un lien ... Alors que pour un système de veille, le veilleur sélectionne les informations pertinentes ce qui offre une réponse claire au système sur les centres d'intérêt du veilleur.

- Le travail en réseau : La communauté de veille permet de définir les tendances globales d'un veilleur comme ses axes de recherche, sa position dans la communauté ... Ce profilage permet alors de construire d'une manière collaborative du sens et offrir une plus-value à l'information traitée (David, 2006).

2.2. Le profil document

Le profil document permet de présenter les aspects génériques ou spécifiques des informations qui peuvent intéresser un veilleur ou une communauté de veille. Il prend en considération trois aspects :

- Les sources d'information : web, flux RSS, base de données... Dans un système de veille, le veilleur ou l'animateur de veille sélectionne les sources d'information qui sont jugées intéressantes et fiables pour un ou plusieurs projets de veille. Ces sources peuvent être formelles (base de brevets, appels d'offre, documents officiels...) ou informelles (information de terrain, forum, émission radio...). On peut aussi distinguer entre les ressources web2.0 (flux RSS, forum, web social, wiki,...) et les ressources du web natif (site web, email, ...) (Bourdier, 2007).

- Le support de l'information: texte, image, vidéo, son..., la gestion de l'information peut varier selon le support sur lequel elle se présente. La prise en considération des annotations, des tags ou des commentaires qui accompagne les objets multimédias et le traitement de texte comme le nettoyage des données superflus, le résumé automatique ou la traduction sont des fonctionnalités offertes généralement par les outils de veille intégrés.

- La thématique de l'information : choisir les thèmes, catégories ou sous catégories susceptibles d'intéresser le veilleur ou la communauté de veille. La

sélection de ces centres intérêts permettra de créer des alertes afin d'avertir les veilleurs sur des nouvelles données qui répondent à leurs problématiques. A ce niveau, il s'agit d'une diffusion, ciblée et automatique des informations dans les réseaux de veille.

Les profils documents sont alimentés au début d'un projet de veille par l'animateur de veille ou le veilleur. Ils seront, ensuite, enrichis d'une manière semi-automatique au cours du cycle du projet par les collaborateurs (communauté de veilleurs).

2.3. Le profil requête

Dans un projet de veille, la définition des requêtes reste l'étape la plus difficile à réaliser par le veilleur, l'animateur ou la communauté de veille. Elle nécessite une connaissance assez élaborée des besoins en veille. Plusieurs auteurs ont proposé des outils d'aide à la formulation de la requête en prenant en considération les centres d'intérêts, les variables booléennes qui permettent d'intégrer un ou plusieurs champs dans la recherche et l'aspect utilisateur (Harbaoui, 2009; Leitzelman, 2009b). Néanmoins, la difficulté reste liée à la compétence du veilleur qui doit être informé en continue de tout changement dans les environnements de l'entreprise. Il doit aussi être à l'écoute de tous ce qui l'entoure (prévoir des réunions de groupes, se déplacer dans les différents services, ...) afin d'éviter le syndrome du veilleur isolé.

Dans le paragraphe suivant, nous présentons notre processus d'intégration d'un système de veille générique qui regroupera ces différents profils dans un objectif de réutilisation optimale.

3. Processus d'intégration d'un processus de veille

Dans ce qui suit, nous allons détailler notre approche d'intégration d'un processus de veille générique à un système de recherche d'information existant en réutilisant les résultats de ce dernier (fichier direct, fichier inverse, noyau d'index...). Notre processus est centré acteur dont l'objectif est d'offrir des informations jugées pertinentes à des utilisateurs cibles. Il est composé de trois phases à savoir : l'intégration, la décomposition et la recomposition de l'information.

3.1. Intégration

Notre point de départ est un système de recherche d'information conçu avec la plateforme open source « Solr » sur un corpus de données bibliographiques sur les nanotechnologies contenant 2893 références bibliographiques complètes regroupant plusieurs domaines de recherche. Nous obtenons ainsi notre matrice terme-document (Noyau0) comme entrée à notre processus de veille.

- La description de la matrice A terme-document :

$$\begin{matrix} & \begin{matrix} I1 & I2 & I3 & \dots & In \end{matrix} \\ \begin{matrix} D1 \\ D2 \\ \dots \\ Dm \end{matrix} & \left(\begin{array}{ccccc} 0.29 & 0.00 & 0.00 & \dots & 0.29 \\ 0.00 & 0.00 & 0.37 & \dots & 0.27 \\ \dots & \dots & \dots & \dots & \dots \\ 0.00 & 0.39 & 0.47 & \dots & 0.25 \end{array} \right) \end{matrix}$$

Avec des occurrences en items (I)

I1: Photonic	D1: id---215
I2: Crystal	D2: id---306
I3: Photonic Crystal	Dm: id---1417
In: Photon	

A la suite, nous procédons à la décomposition en valeurs singulières de cette première matrice, afin d'obtenir des clusters représentatifs des différents environnements couverts par ce corpus.

- La décomposition singulière de la matrice A ($A = U\Sigma V^T$) :

$$S_{i,p} = \text{terme}_i^T * \text{terme}_p \qquad Z_{j,q} = \text{document}_j^T * \text{document}_q$$

$$\left(\begin{array}{cccc} 0.08 & 0 & 0 & 0.08 \\ 0 & 0.15 & 0.18 & 0.10 \\ 0 & 0.18 & 0.35 & 0.21 \\ 0.08 & 0.10 & 0.21 & 0.21 \end{array} \right) \qquad \left(\begin{array}{ccc} 0.14 & 0.07 & 0.08 \\ 0.07 & 0.2 & 0.25 \\ 0.08 & 0.25 & 0.3 \end{array} \right)$$

La matrice U contient les vecteurs propres de S et la matrice V^T contient les vecteurs propres de Z.

$$\widehat{\text{Terme}}^T_i \rightarrow \{[U_1] \dots [U_L]\} \cdot \left\{ \begin{array}{l} \sigma_1 \dots 0 \\ \vdots \\ 0 \dots \sigma_L \end{array} \right\} \left\{ \begin{array}{l} [V_1] \\ \vdots \\ [V_L] \end{array} \right\}$$

Les valeurs singulières ($\sigma_1 \dots \sigma_L$) peuvent alors être sélectionnées à une approximation K afin d'obtenir des clusters à précisions variables et ajustées.

Cette première étape nous a permis d'obtenir les résultats suivants sous forme d'un fichier XML (cf fig1 et tab1).

```

▼<arr name="labels">
  <str>Distortion in Microwave Photonic Filters</str>
</arr>
<double name="score">4.162376843146952</double>
▼<arr name="docs">
  <str>1002</str>
  <str>1030</str>
</arr>
</lst>
▼<lst>
▼<arr name="labels">
  <str>InAs GaAs Quantum Dots</str>
</arr>
<double name="score">20.144567714647327</double>
▼<arr name="docs">
  <str>211</str>
  <str>347</str>
</arr>
</lst>
▼<lst>
▼<arr name="labels">
  <str>One-way Quantum</str>
</arr>
<double name="score">32.09865026310469</double>
▼<arr name="docs">
  <str>1052</str>
  <str>1238</str>
</arr>
</lst>
▼<lst>

```

Figure 1. Fichier XML contenant les différents clusters

Tableau 1. Échantillon de clusters extraits

Cluster	Score	Nbre de doc
Silicon	8,481	96
Resonators	1,974	96
Quantum	6,865	88
Applications	6,455	78
Silicon Photonic	5,139	46
Liquid Crystal	40,5	34
Supercontinuum Generation	21,238	32
Quantum Dots	26,181	33
Photonic Crystal Resonant	5,807	33
Photonic Quantum	14,396	34
Fiber Sensor	2,626	27
Laser Pulses	29,073	18

Pour un ensemble de K traité, nous avons obtenues des résultats de clustering pertinents pour un k=2, générant en totalité 89 clusters. Nous avons créé un nouveau noyau (Noyau 1) sur la plateforme Solr et dans l'objectif est de synthétiser de plus

en plus nos environnements. Nous avons regroupé les clusters obtenus dans des Meta-clusters qui regroupent deux ou plusieurs clusters. Nous avons placé les clusters orphelins dans la catégorie « OtherTopics ».

```

http://localhost:8983/solr/collection1/select?q=%3A*&wt=json

{"responseHeader":{"status":0,"QTime":1,"params":{"wt":"json","q":"","":""},"response":{"numFound":89,"start":0,"docs":[{"id":"1","score":"5,481","cluster":["Silicon"],"IDDoc":["1001","1003","1007","1043","1045","1089","1103","1119","1132","1149","1217","1231","1230","1246","125","1268","1321","133","1335","1352","1358","1379","1384","1397","142","1494","1532","1564","2614","1619","1629","1630","1639","1659","1665","167","1718","1747","194","2027","2038","2039","2144","2050","2055","206","2117","2123","2153","216","2163","2164","2166","2194","2215","222","2226","2254","2260","2266","2289","2294","2299","2322","2328","2331","2335","2352","2436","2437","2475","2482","2507","2526","2539","254","2541","2545","2546","2572","2608","2618","304","305","357","379","386","412","421","435","801","828","929","936","955","973"]},{id":"2","score":"1,974","cluster":["Resonators"],"IDDoc":["1003","101","1025","1027","1037","1052","1083","1089","1113","1134","1154","1157","1166","1211","1214","1217","1234","1247","1261","1271","128","1282","1319","1335","1352","1410","1419","1429","1438","1454","1462","1476","1480","1481","1483","1536","1572","1574","1582","1589","1665","1666","1668","1692","1739","184","194","2030","2061","2112","2189","2190","2202","2208","2238","2243","2244","2284","2352","2353","237","2375","2391","2410","2414","242","2421","2453","2462","2466","2474","2475","2477","2480","2487","2568","2592","261","270","284","311","337","344","375","385","388","391","393","395","858","862","875","880","887","947","992"]},{id":"3","score":"6,865","cluster":["Quantum"],"IDDoc":["1016","1044","1045","1052","1056","1098","1134","1140","1161","1165","1176","1179","121","122","1227","1238","1252","1261","1277","1280","1282","1287","129","1334","1362","1393","1413","1422","1441","1536","155","1582","1584","1588","1608","1660","1679","1684","1699","1710","2020","2060","2062","2066","2107","211","2174","2245","2259","229","2366","2370","2395","2418","2421","2448","2497","2499","2519","253","2548","2568","2569","257","293","341","347","349","356","385","392","423","432","842","890","899","902","905","906","910","948","949","953","961","964","983","985","986"]},{id":"4","score":"6,455","cluster":["Applications"],"IDDoc":

```

Figure 2. Fichier Json contenant les Meta-Clusters

Le résultat obtenu est sous forme d'un fichier Json contenant les Meta-Clusters. Les clusters regroupent ainsi les identifiants des documents qu'ils représentent. Ces résultats seront enregistrés dans un troisième noyau (Noyau 2) Solr afin de faciliter la recherche dans ces Meta-Clusters.

La deuxième phase de notre processus de veille générique présente la décomposition de l'information.

3.2. Décomposition

L'objectif de cette deuxième phase est de décomposer les environnements de veille sur les différents veilleurs (ou communautés de veille) selon leurs centres d'intérêt. Afin de tester notre application nous avons construit quatre profils de veilleur qui représentent des environnements différents.

- Profil requête : à la différence d'un système de recherche d'information, le veilleur va sélectionner les documents qu'il va juger pertinent. La matrice requêtes/documents est alors une matrice binaire (1 pour les documents sélectionnés et 0 pour les documents ignorés)

$$\begin{matrix}
 & Rq_1 & Rq_2 & \dots & Rq_n \\
 \left. \begin{matrix} D_1 \\ D_2 \\ \dots \\ D_m \end{matrix} \right\} & \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ \dots & & \\ 0 & 1 & 1 \end{pmatrix}
 \end{matrix}$$

- Centres d'intérêts : le veilleur sélectionne ses centres d'intérêts en amont du processus de veille. La matrice centre d'intérêt/veilleur est aussi une matrice binaire (1 pour les centres d'intérêts sélectionnés et 0 pour ceux ignorés).

$$\begin{matrix}
 V_1 \\
 V_2 \\
 \dots \\
 V_k
 \end{matrix}
 \left\{
 \begin{array}{ccc}
 CI_1 & CI_2 & CI_x \\
 1 & 0 & 0 \\
 0 & 1 & 1 \\
 \dots & & \\
 1 & 1 & 0
 \end{array}
 \right\}$$

- Profil document : le profil document permet de spécifier les préférences du veilleur concernant les résultats retournés.

La première étape est la classification de tous ces champs représentatifs du veilleur dans une base de connaissances spécifique à chaque veilleur (cf. Fig3).

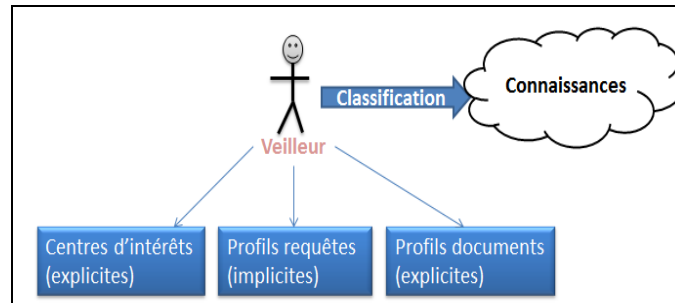


Figure 3. Projection du veilleur sur l'axe des connaissances

La deuxième étape consiste en un « matching » supervisé des différents veilleurs par rapport au Meta-clusters extraits dans la première phase du processus (intégration) : (cf. Fig4)

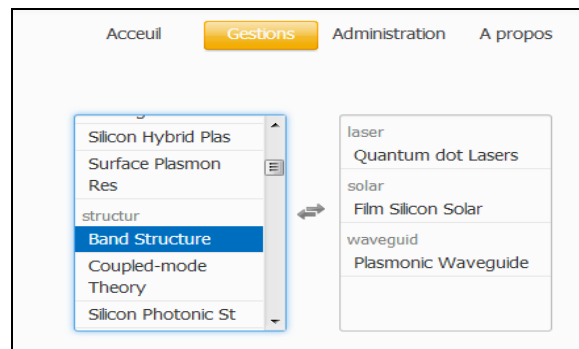


Figure 4. Interface sélection des Meta-Clusters

Le résultat de cette deuxième phase est une matrice de décomposition $Decomp\{MetaClust\}.\{Veilleur\}$. La dernière phase la recombposition sera présentée dans le paragraphe suivant.

3.3. Recomposition

La décomposition des informations selon les préférences implicites et explicites de chaque veilleur a permis de mieux viser les besoins de ce dernier afin de remédier au problème majeur des systèmes de veille qu'est la surcharge cognitive. Néanmoins, le résultat final attendu par le décideur ou la communauté d'experts doit regrouper toutes les briques de résultats retournés par les veilleurs. Cette dernière phase consiste alors à regrouper les résultats qui sont présentés sous forme de documents sélectionnés, de les présenter sous forme d'informations à valeur ajoutée et de les diffuser dans les réseaux concernés. Chaque veilleur sélectionne les articles qu'il juge pertinents (cf Fig.5). Ensuite, il les diffuse sur le réseau.



Figure 5. Plateforme principale Gwatch

Le regroupement des résultats obtenus est une matrice binaire document/ veilleur où on attribue la valeur 1 à chaque document i sélectionné par un veilleur j et 0 sinon.

$$\begin{matrix}
 & \left. \begin{matrix} D1 & D2 & Dm \end{matrix} \right\} \\
 \begin{matrix} V1 \\ V2 \\ \dots \\ Vk \end{matrix} & \left\{ \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ \dots & & \\ 0 & 0 & 1 \end{matrix} \right.
 \end{matrix}$$

La hiérarchisation des informations a permis de prendre en considération deux aspects : la pertinence de l'information et la position du veilleur dans le réseau de veille. En illustration sur un exemple :

Veilleur	H
V1	1
V2	2
V3	2
V4	3

Document	H
D1	1
D2	2
D3	3

Le processus de veille proposera aux veilleurs, les résultats suivants:

V1{D1,D2,D3}; V2{D1,D2};
V3{D2,D3}; V4{D1}

4. Résultats et discussion

Un système de veille, étant un système d'information (SI), sa mise en place comporte comme tout SI quatre phases successives à savoir : l'étude, le développement, l'implémentation et la clôture (Caron-Fasan, 2010; Schwalbe, 2004). La phase clôture ou bilan permet d'évaluer l'apport d'un tel système ainsi que ses limites.

Notre système de veille générique, dans son état actuel, est un prototype complexe qui a été simplifié, qui vise essentiellement à faciliter la mise en place d'un système de veille tout en prenant en considération les besoins spécifiques du veilleur et les différentes facettes d'un projet de veille. Notre prototype ne contient pas toutes les fonctionnalités d'un système de veille intégré classique (résumé automatique, traduction, gestion des flux RSS...). Néanmoins, il apporte une réponse claire aux deux problèmes majeurs des systèmes sur le marché, à savoir la résolution de la surcharge cognitive et la perte dans l'hypermédia. La centralisation sur le profil acteur, ses besoins, ses projets antérieurs et sa position dans la communauté de veille, nous ont permis d'obtenir des interfaces spécifiques pour chaque acteur et de viser l'information cible selon ses besoins.

La construction de notre système de recherche d'information par la plateforme Solr, sur le corpus de données bibliographiques en nanotechnologie, nous a permis d'obtenir, comme entrée à notre système de veille générique, le fichier direct et le fichier inverse (CF Fig.6) correspondant au corpus d'analyse.

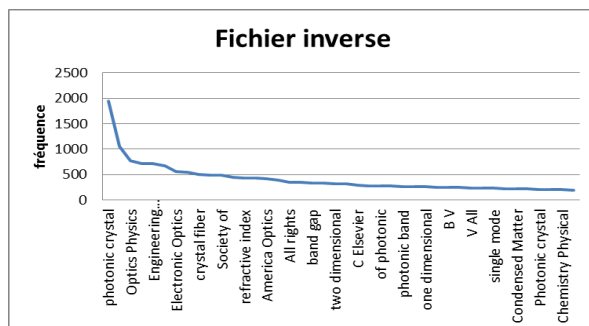


Figure 6. Fichier inverse index/fréquence (loi de Zipf)

Nous avons construit grâce à la méthode de décomposition en valeurs singulières par un algorithme de clustering non supervisé Lingo qui est implémenté dans la plateforme Solr par l'API carrot2[21]. Nous avons obtenu pour la valeur arbitraire $K=2$ le nombre de 89 clusters (voir figure suivante) avec un degré de pertinence. Un nombre de documents n'ont pas été classés et ils ont été placés dans un cluster général (divers) (cf. Fig7).

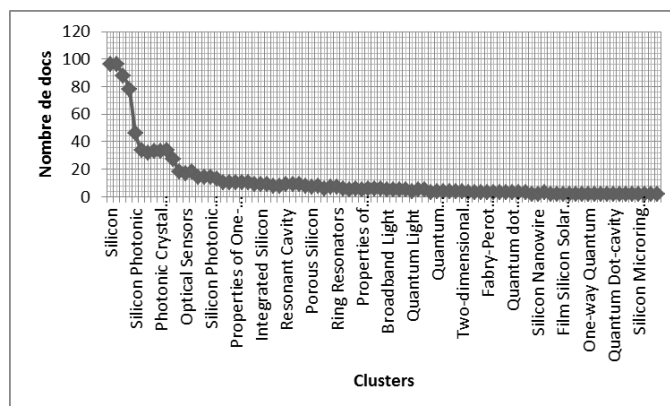


Figure 7. Courbe des clusters/nombres de documents

Dans l'objectif de synthétiser nos résultats, nous avons regroupé ses différents clusters avec un algorithme de clustering non supervisé dans des méta-clusters. La figure 8 schématise le nombre de cluster par Meta-cluster tout en sachant qu'un méta-cluster regroupe au moins 2 clusters et qu'un cluster peut se trouver dans un ou plusieurs Meta-clusters.

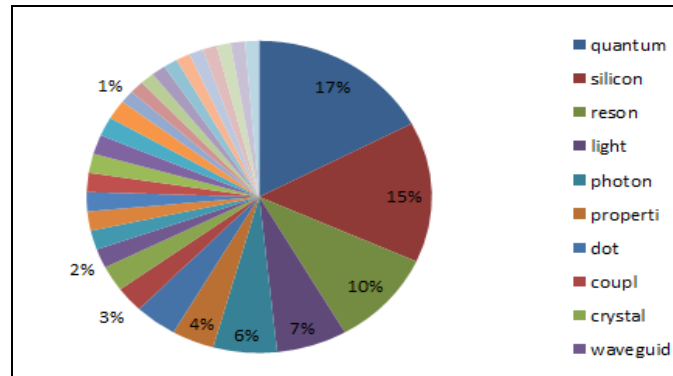


Figure 8. Classification des clusters dans des Meta-clusters

Dans le contexte de veille, l'évaluation du système proposé ne répond pas aux mêmes critères (calcul de précision et de rappel..) qu'un système de recherche d'information (Baccini, 2008) même si ce dernier est considéré comme un noyau pour un système de veille et donc sa pertinence est très importante pour le retour d'informations recherchées. Des évaluations ont été proposées, comme à titre d'exemple, les phases de veille couvertes, les outils et fonctionnalités offerts, les sources et types d'informations gérés... D'un autre côté, on peut proposer des évaluations sur l'impact d'un tel projet dans une entreprise ou une organisation: l'implication des personnels, la pertinence des retours, l'impact sur le processus décisionnel ... (Jakobiak, 2002).

A cet effet, nous proposons un tableau montrant nos apports par rapport aux différentes phases d'un projet de veille à savoir : définition des besoins, recherche, collecte, traitement et diffusion (cf Tab.2)

Tableau 2. Les apports de notre approche

Phase	Apports
Définition des besoins	Faciliter la définition des besoins grâce à un palmarès de Meta-clusters et clusters afin de mieux se situer par rapport au corpus.
Recherche	Précision dans la recherche par la sélection préalable de la plage des documents susceptibles d'être pertinent pour le veilleur.
Collecte	La sélection directe des documents jugés pertinents par le veilleur et la constitution d'un panier de documents qui répond à la problématique de veille.
Traitement	La possibilité de nettoyer les documents des informations jugées superflues.

Phase	Apports
Diffusion	Le regroupement et la hiérarchisation des résultats permettent d'obtenir un retour plus pertinent pour l'utilisateur final (décideur, expert...). La sécurité de l'information est assurée par la sélection au préalable par le veilleur ou l'administrateur de la communauté qui peut accéder à ces résultats.

Les limites de ce travail de recherche restent dans la possibilité de développer un système de veille complet ce qui nécessite un travail d'équipe de développeurs et plusieurs mois de développement. Néanmoins, nous considérons que notre prototype est assez représentatif de notre approche. Pour compléter notre phase test avec l'aspect utilisation, nous avons besoin d'une validation par une équipe de veilleurs spécialisés en Nanotechnologie, dans plusieurs catégories et la validation de plusieurs projets distincts.

Nous terminons cet article par notre conclusion et la présentation de nos perspectives.

5. Conclusion et perspectives

Le contexte de la recherche, à savoir les systèmes de veille, nous a permis d'identifier, d'un côté, que le noyau de tout système de veille est un SRI qui exécute des requêtes et permet l'indexation des informations collectées, et d'un autre côté, la place centrale du veilleur dans le système. Ces constats nous ont permis de nous orienter vers un système de veille générique centrée acteur et de coupler un système de recherche d'information existant, pour réutiliser ses résultats (fichier directe et inverse) et ses composants (noyau d'indexation).

Notre proposition a porté sur une approche comprenant trois phases, à savoir :

- l'intégration des informations par un processus non supervisé de clustering et l'identification des Meta-clusters ;
- La décomposition des informations selon les besoins de chaque veilleur ;
- La recomposition des résultats dont l'objectif est de fournir une réponse complète à la problématique de veille.

L'application de cette approche sur un corpus de données bibliographiques sur la Nanotechnologie comprenant 2893 références, nous a permis d'identifier 89 clusters répartis sur 34 Meta-Clusters. Nous avons ensuite décomposé un nombre de ces Meta-clusters sur quatre profils veilleur. Notre prototype « GWatch » développé a permis alors de répondre à des besoins spécifiques en offrant des informations collectées à des veilleurs ciblés.

Dans nos perspectives, nos futures orientations consistent essentiellement à pouvoir enrichir notre prototype par les différentes fonctionnalités qu'offre un système de veille intégré (Gestion des flux RSS, résumé automatique, traduction, surbrillance des mots clés dans les articles...) et sa validation par une équipe d'experts dans le domaine de la Nanotechnologie.

Bibliographie

- Baccini, A., Déjean, S., Kompaoré, D., Mothe, J. (2008). Analyse des critères d'évaluation des systèmes de recherche d'information. *ISI_BDKM*.
- Bernat, J.-P., al. (2008). Les contours de la veille *Documentaliste-Sciences de l'Information* 45. 32-44
- Bourdier, S. (2007). Enjeux et apports du web 2.0 pour la circulation de l'information dans l'entreprise: le cas du service de veille stratégique du groupe Yves Rocher. *Institut national des techniques de la documentation*.
- Caron-Fasan, M.L., Lesca, H. (2010). Facteurs de risque de la conduite de projet de mise en place d'un dispositif de veille anticipative dans plusieurs Caisses d'Allocations Familiales *Colloque de l'AIM, La Rochelle*. 20 pages
- David, A. (2006). La recherche collaborative d'information dans un contexte d'intelligence économique. *Le système d'information de l'entreprise*.
- Delaby, A. (2004) "Le concept de veille", UFR Tours -Ecole Doctorale
- Fayard, P.M. (2002). "Le concept de "Ba" dans la voie japonaise de la création du savoir". *Ambassade de France à Tokyo : Service pour la science et la technologie* SMM03-046.
- Harbaoui, A., M. Ghenima, S. Sidhom. (2009). Enrichissement des contenus par la réindexation des usagers: un état de l'art sur la problématique. *SIIE*.
- Jakobiak, F. (2002). Evaluation de la veille technologique. *Intelligencia competitiva*.
- Jakobiak, F. (2005). *De l'idée au produit: Veille-R&D-Marché*. Editions d'organisateur
- Kislin, Phil. (2007) "Modélisation du problème informationnel du veilleur dans la démarche d'intelligence économique" in *Sciences de l'Information et de la Communication*. Nancy, Université Nancy 2
- Leitzelman, M. (2009a). Etat de l'art et tendances sur le marché de la veille et l'intelligence compétitive. *ISICIL*.
- Leitzelman, M., Ereteo, G., Grohan, P., Herledan, F., Gandon, F., Buffa, M. (2009b). De l'utilité d'un outil de veille d'entreprise de seconde génération. *IC*.
- Lesca, H. (2001). Veille stratégique : passage de la notion de signal faible à la notion de signe d'alerte précoce. *VSSST*.
- Lesca, N. (2011). Etat des lieux des pratiques de « veille logistique durable » : une approche qualitative, Rapport. *ANNEXE N°3 au Rapport scientifique d'étape VLD.1-PREDIT 4-ADEME*.
- Pateryon, E. (1998). La veille stratégique. *Economica*.
- Raquel, J.M., Fabio D.B., Lesca H., Freitas H. (2011). Application de la veille anticipative stratégique pour le suivi de l'environnement et la production de connaissances actionnables *Journal of Information Systems and Technology Management* 8. 425-440
- Schwalbe, K. (2004). *Information technology project management*. 3 ed.
- Sidhom, S. (2002) "Plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information : de l'écrit vers la gestion des connaissances". France, Université Claude Bernard - Lyon 1
- Sidhom, S., P. Lambert. (2011). Information Design for weak signals detection in Economic intelligence. *SIIE*. 123
- Thiery, Odile, David, Amos. (2002). Modélisation de l'utilisateur, Systèmes d'Informations Stratégiques et Intelligence Economique. *Revue Association pour le Développement du Logiciel (ADELI)* N 47. 12 p
- Thomas, A. (2008). Parce que la veille bouge. *Documentaliste-Sciences de l'Information* 45. p.30-31