RECHERCHE DE COMBINAISONS DE WEB SERVICES 2.0 POUR OUTILLER LE VEILLEUR

Mylène LEITZELMAN (*)
Mylene.Leitzelman@telecom-paristech.fr

(*) Telecom Paristech, 2229 route des crêtes, BP 193. 06560 Sophia-Antipolis

Mots clefs :
Mashup, APIs, veille 2.0, innovation, analyse des réseaux sociaux
Keywords:
Mashup, APIs, enterprise 2.0, innovation, social network analysis
Palabras clave :
Mashups, APIs, inteligencia 2,0, innovación, análisis de redes sociales

Résumé

Le terme « mashup » désigne une application combinant des fonctionnalités ou du contenu en provenance de différentes sources d'information ou de sites web ; son évolution est donc fortement corrélée à l'explosion des applications de création, d'expression et d'échange, de partage et de réseaux sociaux caractérisant le web 2.0 qui recentre beaucoup le pouvoir au main du consom'acteur (cf. syndrome du self-service et du DIY – Do It Yourself). Pour preuve, dans leurs pronostics sur l'avenir du marché de la Business Intelligence, les analystes d'instituts privés comme IDC [1], Forrester [2] ou Gartner [3] sont unanimes sur la nécessité de donner plus de pouvoir de décision à l'utilisateur final afin qu'il puisse manipuler et combiner ses propres vues de données internes et externes à l'entreprise dans un environnement de fonctionnalités type self-service. Nous pensons que ces évolutions ne sont pas sans conséquence aussi sur les solutions classiques de gestion de l'information et des connaissances car les mashups peuvent proposer une alternative intéressante pour outiller les besoins du veilleur en matière de collecte, analyse, partage et diffusion de données actionnables.
Il s'agit donc dans ce présent article d'une analyse exploratoire dont l'objectif est de trouver et sélectionner des combinaisons de web services pour outiller le veilleur dans son travail quotidien en étudiant notamment les applications actuelles de mashups et d'APIs proposées d'une part dans la littérature scientifique et d'autre part dans le répertoire de référence Programmableweb. Pour ce faire, nous avons utilisé deux méthodes d'analyse, l'une bibliométrique, l'autre d'analyse des réseaux sociaux sur les matrices mashups/APIs obtenues. Finalement, nous avons exploité la complémentarité des deux méthodes pour mettre en lumière un certain nombre de services web composés de mashups et d'APIs que le veilleur peut mobiliser en contexte de veille.

PLAN

RECHERCHE DE COMBINAISONS DE WEB SERVICES 2.0 POUR OUTILLER LE VEILLEUR 1
Résumé 1
1 Introduction : notre problématique, recenser les mashups et APIs pour la veille 3
2 Les mashups : revue de positionnement et classification 3
2.1 Définition 3
2.2 Classifications des mashups 5
2.3 Les mashups pour la veille ? 5
3 Les mashups dans la littérature scientifique : une recherché tirée par les leaders du marché 7
3.1 Méthode de recherche : analyse bibliométrique du corpus 7
3.2 Premiers résultats : tendances fortes sur les « Next-Generation Mashup » 10
4 Le marché des mashups : analyse du répertoire en ligne ProgrammableWeb 12
4.1 Méthodologie adoptée : l'analyse des réseaux sociaux pour faire émerger des combinaisons singulières 12
4.1.1 Premières révélations sur la structure globale du réseau : les mashups incontournables pour la veille 13
4.1.2 Repérage d'éléments singuliers au sein des réseaux 16
4.2 Nos résultats de combinaisons d'APIs et de mashups pour la veille 17
5 Conclusion et perspectives 22
5.1 Les mashups : un marché pour la veille 22
5.2 Perspectives de recherche et approfondissements 22
6 Bibliographie 23
7 Annexes 26
7.1 Tagclouds des sociétés publiantes (extraction des mots-clés du champ KW) 26
7.2 Liste des conférences 2010-2009 27
7.3 Codes de classification d'INPEC dédiés à la veille 28
7.4 Visualisation des clusters par famille de mashups autour des tags représentatifs de la veille 29
7.5 Réseau complet des combinaisons de mashups et APIs pour la veille 30

1 Introduction : notre problématique, recenser les mashups et APIs pour la veille

Cet article constitue un des livrables du projet ISICIL - Intégration Sémantique de l'Information par des Communautés d'Intelligence en Ligne - qui a pour but de concevoir, étudier et expérimenter l'utilisation de nouveaux outils d'assistance aux tâches d'intelligence collective en entreprise en s'appuyant sur les interfaces avancées des applications du web 2.0 pour les interactions et sur les technologies du Web Sémantique pour l'interopérabilité et le traitement de l'information. Suite à un état de l'art et des tendances du marché de la veille (2.0) et de l'Intelligence compétitive réalisé en 2009[1], il s'agissait ensuite d'effectuer un travail de sélection sur internet de mashups, d'APIs et autres web services pouvant être exploités en contextes de veille, l'objectif étant au final d'une part de surveiller les initiatives concourantes voire concurrentes du projet ISICIL et d'autre part d'envisager l'intégration de certaines applications détectées dans les prototypes développés. Nous présentons ci-après la démarche adoptée et les premières analyses obtenues pour répondre à ces objectifs. Nous avons opté pour une démarche exploratoire, d'une part en analysant comment les mashups sont perçus et avec quoi ils sont le plus associés dans la littérature scientifique (base INSPEC[2]) et ensuite par l'analyse en profondeur du répertoire de mashups ProgrammableWeb, rare référence fiable actuellement dans ce domaine[3]. Nous proposons tout d'abord une revue de définition et classification des mashups et leur utilisation éventuelle pour la veille (chapitre 2). Nous présentons ensuite notre analyse du positionnement de la communauté scientifique sur la question des mashups, où des tendances de fond et d'avenir se dégagent clairement (chapitre 3). En complément de cette démarche, la partie suivante est consacrée à l'analyse du marché des mashups sous le prisme du répertoire en ligne ProgrammableWeb proposant plus de 5000 mashups et 2000 APIs (chapitre 4) ; nous verrons notamment dans cette partie comment, à l'aide de métriques issues de l'Analyse des Réseaux Sociaux, nous pouvons effectivement mettre en lumière des combinaisons de mashups et d'APIs incontournables et singulières propres à aider le veilleur dans son travail d'analyse. Au chapitre 5, nous concluons sur l'avenir des mashups pour la veille et les perspectives de recherche que nous offrent cette première étude exploratoire.

2 Les mashups : revue de positionnement et classification

2.1 Définition

Si le terme « mashup » apparaît d'abord dans le milieu musical entendu comme mix de plusieurs bandes son pour créer de nouvelles musiques [4], ce terme est plutôt récent dans l'ère numérique (premières apparitions en 2005 coïncidant notamment avec la mise en ligne de Google Maps [5]) et dont on peut considérer sa dynamique comme fortement corrélée à l'explosion des applications sociales et virales caractérisant le web 2.0.

Un mashup désigne une application dite composite qui permet de combiner des données, des fonctionnalités ou des services provenant de plusieurs applications tierces (ie *third party*) pour finalement créer une nouvelle application, originale et distincte de ses éléments d'origine. Dans le cas de services Web, le principe d'un mashup est d'agréger des fonctionnalités et des contenus provenant de fournisseurs de ressources qui facilitent l'accès à leurs données via une API (i.e. *Application Programming Interfaces*), ce terme désignant souvent indistinctement tant une documentation technique que le nom de son fournisseur. Grâce aux technologies type du web 2.0 (comme REST, SOAP, AJAX,…) il est devenu beaucoup plus facile pour les utilisateurs internautes de combiner des bribes de chansons, de vidéos ou de textes à partir de différentes sources pour créer du contenu nouveau et intéressant : « le mashup constitue donc une forme d'externalisation qui repose sur une concentration des efforts sur l'expérience utilisateur plutôt que sur ses capacités propres de développement »[4].

---

[1] Voir http://isicil.inria.fr/docs/Livrables/ISICIL-ANR-EA02-BusinessIntelligence-0906.pdf

[2] Base de données dédiée au STIC avec plus de 11 millions d'articles scientifiques, actes de colloques et rapports techniques indexés (via la banque de données EBSCO de l'INIST), date de l'extraction : avril 2010

[3] Quelques études scientifiques ont aussi porté sur l'analyse de sa base de données, depuis 2007 et dont nous nous inspirerons pour la suite de notre étude [41] [49]

[4] Cf. Qu'est ce qu'un Mashup via API Orange : http://api.orange.com/wiki/index.php/Qu%27est-ce_qu%27un_mashup

Sur Internet, l'intérêt sur les mashups provient essentiellement du monde de l'entreprise (tendance qui se confirme aussi dans l'analyse de la littérature scientifique, cf. Chapitre 3), et plus précisément nous retrouvons de manière récurrente de grandes sociétés fournisseurs de contenu et de technologies web (comme IBM, Serena, Microsoft, Google) et quelques startups très tendances (comme OpenKapow ou Jackbe). En effet, depuis ces cinq dernières années, l'écosystème des mashups est dominé par les grandes majors de l'internet (i.e. Google, Yahoo! ou Amazon) qui proposent l'accès libre à leurs données pour être remixées et reformatées sans fin par les internautes [6][5]. L'utilisation de l'option GoogleTrends permettant de visualiser l'intérêt des recherches effectuées sur le terme « mashup », avec une répartition temporelle comme géographique, vient confirmer cette prédominance des acteurs économiques sur le marché des mashups.

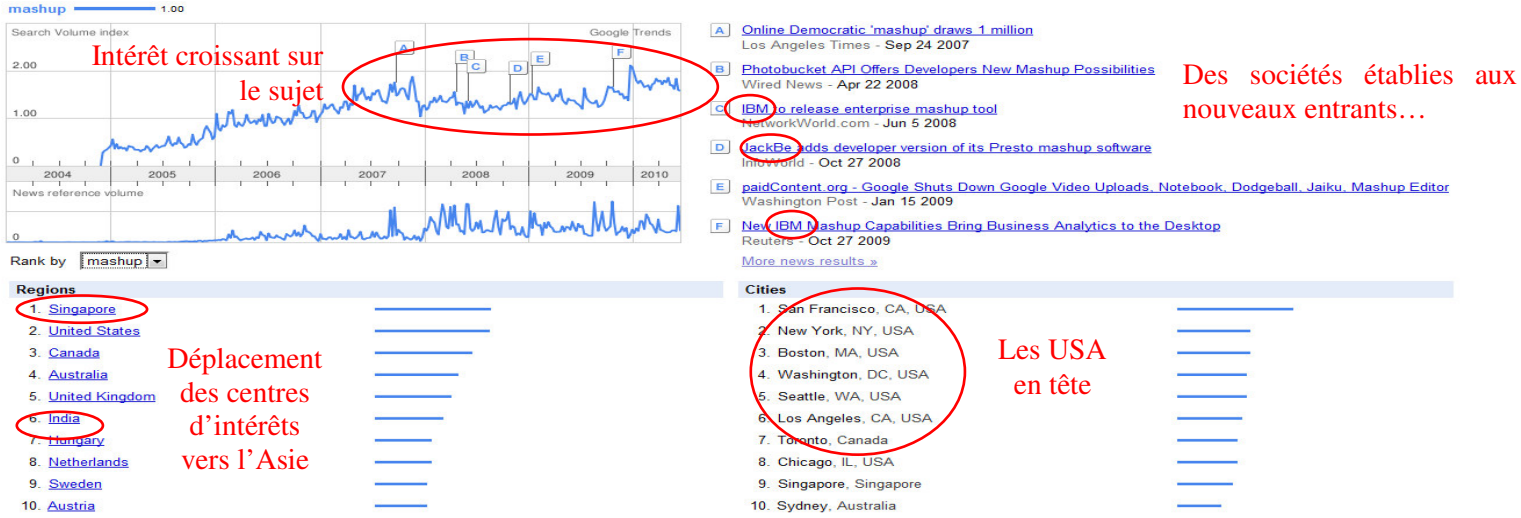


Figure 1 L'option GoogleTrends pour les recherches effectuées sur "mashup"

Il est intéressant de voir Figure 1 **la place des USA** dans le top des régions la présence des villes comme San Francisco, Los Angeles, New-York et Boston, qui ne sont autre que les villes d'implantation des principales startups du domaine comme OpenKapow, Jackbe ou Serena. Dans le top 10 des pays, nous trouvons aussi la présence de l'Inde et de Singapour comme forces de recherche montantes dans le domaine des mashups, éléments confirmés par la suite via l'analyse du corpus d'INSPEC. Nous attirons aussi l'attention sur la non présence significative de la Chine dans le moteur de Google, n'indexant pas la production chinoise, qui par contre est très présente dans la littérature scientifique.

[5] Côté public, nous retrouvons aussi un mouvement qui émerge en force pour la libération des données publiques (ex via l'initiative d'Obama aux USA http://www.whitehouse.gov/Open, ou avec le mouvement du *linked data*, issu du web sémantique pour lier les données ouvertes entre elles http://linkeddata.org/)

2.2 Classifications des mashups

Les mashups peuvent s'appréhender de différentes manière, et dans la littérature en ligne lui étant consacrée, surtout par l'intermédiaire des nombreux livres blancs ([7], [8]), trois principales classifications de mashups émergent : la première centrée sur le type de données mixées, la seconde orientée sous un angle plutôt fonctionnel et la troisième axée sur la chaine de valeur des technologies impliquées[6].

En terme de classification par le contenu, nous avons quatre grandes catégories de mashups : les mashups de cartographie et géolocalisation de données, présentant toutes sortes de données sous forme de cartes géographiques essentiellement via les APIs de Google Maps ou Yahoo Maps, les mashups de contenu multimédia combinant les métadonnées provenant des vidéos, photos ou images, les mashups de recherche d'information et d'e-commerce, centrés sur la comparaison de prix et de produits (dont l'explosion est fortement liée aux initiatives d'Ebay et d'Amazon d'ouvrir l'accès à leurs bases) et enfin les mashups de news, rendus possible via l'agrégation de flux RSS/Atom pour la création personnalisée de pages d'accueil ou de journaux en ligne.

Une autre taxonomie, issue d'un récent rapport de l'institut Forrester [2] propose une catégorisation fonctionnelle des mashups suivant leur degré de sophistication : les mashups les moins sophistiqués mais les plus répandus forment la catégorie des mashups de présentation, autrement appelés « mashups grand public » (i.e. *Consumer Mashup*) car ils sont créés par l'utilisateur pour répondre à un besoin précis et contextuel ; ces derniers sont formés par la combinaison de diverses données et d'éléments visuels agrégés pour former une vue unifiée (cf. pour meilleur exemple le portail de flux RSS Netvibes ou encore le service Yahoo! Pipes [9]). Avec un degré de sophistication plus important, nous avons la catégorie des Mashups de données, unifiant sous une vue unique plusieurs fils de contenu (cf. par exemple le site TwitterVision d'agrégation de fils d'informations postés dans Twitter sous une vue cartographique avec l'API de Google Maps). Les mashups de process ou de business forment la catégorie de mashups la plus sophistiquée, car elle implique non seulement le mixage de données business, mais aussi de processus et de modèles de business avec une notion importante de personnalisation et de partage au sein de l'entreprise. C'est sur cette dernière catégorie que se fondent les espoirs comme les craintes sur l'essor du marché des mashups et notamment dans le monde de l'entreprise, nous y reviendrons par la suite.

Enfin, un dernière classification présente les mashups sous l'angle des technologies en présence dans la chaine de valeur du marché : d'un côté il y a les fournisseurs de contenu et d'APIs qui produisent les données sources ; les plus populaires étant les services de géolocalisation comme Google Maps largement en tête, Yahoo Maps et Microsoft Virtual Earth, les services de partage de média et de réseaux sociaux comme Flickr, YouTube, Del.icio.us, Twitter, Facebook, Last.fm ou encore les services de recherche d'informations comme Google Search, Yahoo Search, ou dédiés à l'e-commerce comme Amazon ou eBay. Ensuite dans la classification se trouvent les sites de mashups hébergeant les applications utilisatrices de données des fournisseurs de contenu et qui peuvent être exécutées soit directement sur le client (via des technologies comme Ajax, javascript, XHTML,…) soit dans une logique de génération dynamique de contenu côté serveur (via des Java Servlets, scripts CGI, PHP,…). Finalement en bout de chaîne se trouve le navigateur client, par lequel se passe l'essentiel de l'expérience utilisateur de la manipulation des mashups, ce qui explique leur grande popularité des mashups auprès du public et qui nous pousse à explorer en quoi ils peuvent outiller le veilleur dans son travail de recherche, d'analyse, de mise en forme et de diffusion d'informations pour la prise de décision.

2.3 Les mashups pour la veille ?

Tout d'abord, les recherches centrées sur les implications des Technologies de l'Information et de la Communication et notamment l'utilisation d'internet dans le processus de veille ou d'intelligence compétitive ne sont pas nouvelles et ont attiré l'attention de nombreux chercheurs depuis une quinzaine d'années (voir [10] pour une revue théorique complète de littérature sur ce sujet ou [11] pour une recherche empirique sur plus de 500 entreprises européennes). Plus récemment, citons l'un des derniers Cahiers du Numérique « Du web 2.0 au concept 2.0 »[7] dont les articles dissèquent les conséquences du web 2.0 sur les pratiques du veilleur qu'il

[6] Pour une revue approfondie des diverses classifications de mashups, voir [47] proposant suite à un modèle de classification un guide méthodologique de construction de mashups, ou encore [48] pour des travaux similaires sur les mashups d'entreprise.

[7] Voir les articles du numéro LCN consacré à la veille et au web 2.0 : http://www.cairn.info/revue-les-cahiers-du-numerique-2010-1.htm

soit documentaliste, journaliste ou travailleur du savoir en entreprise, et introduisant le concept de veille 2.0, où le veilleur en utilisant les technologies sociales, virales et faciles d'accès du web 2.0 devient l'acteur principal de bout en bout du processus de veille [12].

Du côté du marketing, de récents rapports provenant d'instituts privés de recherche comme le Forrester Institute [13] et le Gartner Group [3] font également état d'une évolution marquée de tout un pan de la Business Intelligence vers l'intégration de mashups dans leurs applications (d'après Gartner, 1/3 des applications analytiques de BI seront sous forme de Mashups à l'horizon 2013[8]). Ils voient en effet dans les mashups un moyen de démocratiser l'accès aux applications analytiques de collecte, traitement et diffusion d'informations, ouvrant ainsi la porte du sérail aux « employés du savoir » (ie *white collars*) et aux décisionnaires, monde jusqu'à peu réservé aux seuls analystes et départements informatiques. Les avantages de proposer en libre service des applications de recherche de données, interopérables entre les données internes et celles disponibles sur Internet, facilement agrégeables via les standarts RSS et XML, filtrées aux moyens de visualisations multiples, sous forme de dashboard ou de reporting auto-alimentés, contextualisés et customisés, offrant la possibilité de partager des vues, des créations, des commentaires,... le tout avec son réseau social permettent ainsi de déplacer la création de services vers l'utilisateur final et de réduire ainsi le délai entre le besoin d'information et la prise de décision. Dans un registre plus global de gestion des connaissances, Etienne Wenger voit aussi clairement le potentiel énorme des web services 2.0 pour outiller ces communautés dans son récent livre dédié à l'analyse des technologies pour manager les communautés de pratique [9].

Enfin, nous sommes convaincus que les mashups de données, de présentation et de processus peuvent répondre à chaque besoin fonctionnel du cycle de la veille. Pour étayer cette assertion, nous sommes partis des travaux de recherche issus de l'analyse multi-points de vue des besoins de communautés de veilleurs réalisés dans le cadre du projet ANR ISICIL[10]. L'analyse des scénarios d'usage des terrains d'expérimentation du projet ISICIL, corrélée aux fonctionnalités de solutions issues du marché[11] a résulté, entre autre, à la mise en forme d'une ontologie heuristique des fonctions de veille articulées autour de cinq blocs fonctionnels dans lesquels ont été réparties des fonctions de Système d'Information propres à équiper le veilleur[12]. C'est aussi sur la base de cette ontologie de veille que nous avons sélectionné nos ensembles de mots-clés (descripteurs ou tags) pour cibler au mieux l'analyse de nos corpus sur les mashups dédiés à la veille dans les chapitres suivants.

Nous présentons ci-après notre méthodologie d'investigation de deux corpus de données dédiés aux mashups, l'un provenant de la base de données scientifique INSPEC, l'autre issu du répertoire en ligne de ProgrammableWeb pour une analyse empirique du marché des mashups.

---

[8] Voir http://www.gartner.com/it/page.jsp?id=856714

[9] "*We see communities combining tools, ready to move across and between platforms to find tools with the features they need for their current purposes. Opportunistic and cost-sensitive, they continue to actively "configure" a set of technology resources that fit their changing understanding of their changing needs... From a bridging perspective, it may be more productive to have light modules that are made compatible through standards like XML, API, and RSS syndication. These modules may be single tools or small clusters of neighbouring tools that can be "plugged" into each other or "talk" to each other.*" [46]

[10] Voir http://isicil.inria.fr/res/docs/livrables/ISICIL-ANR-EA02-BusinessIntelligence-0906.pdf

[11] Il s'agit plus précisément du marché des outils de Business Intelligence, celui des outils de veille intégrant tout ou partie des étapes du cycle de l'information et celui des outils sociaux du web 2.0 dont les tendances sont détaillées dans le livrable [51]

[12] Voir la carte heuristique des fonctionnalités de veille : http://isicil.polytech.unice.fr/maps/fonctions_veille.html et en Annexes 8.5 la représentation de ces blocs enrichis d'exemples de mashups détectés au cours de nos analyses

3 Les mashups dans la littérature scientifique : une recherché tirée par les leaders du marché

3.1 Méthode de recherche : analyse bibliométrique du corpus

Nous avons choisi d'interroger la base de données INSPEC[13] sur le terme générique « mashup » en multi-champs, afin d'une part de nous faire une idée sur le positionnement de la communauté scientifique sur ce terme et, d'autre part, de mettre en lumière des applications effectives de mashups en contexte de veille. Nous avons obtenu un corpus constitué de 327 références **majoritairement récents** (40% de 2009, 38% de 2008, 18% 2007, 2% 2006 et seulement 7 articles pour 2010), ce qui se distingue des dates plus anciennes trouvées via le moteur Google sur Internet. Nous avons traité bibliométriquement les champs de notre corpus d'INSPEC, pour obtenir globalement 788 auteurs distincts, 2035 mots-clés (des auteurs), 301 termes index (du thésaurus d'INSPEC), 68 sources de journaux et 163 conférences internationales.

Plusieurs éléments d'analyse du corpus nous permettent d'avancer que la recherche sur les mashups est une recherche très appliquée, principalement alimentée par les grands fournisseurs de solutions et leurs centres de R&D. Nous détaillons ci-après les points clés de notre analyse :

- **Une recherche récente très collaborative mais sans réel noyau dur** : nous retrouvons cette caractéristique de recherche récente dans le fait que 86% des auteurs identifiés (soit 684 sur un total de 789 auteurs) n'apparaissent qu'une fois dans le corpus, 10% avec 2 références et seulement 2% avec plus de 3 références. Par contre, il est à noter que 23% des articles n'ont qu'un seul auteur, alors que 44% sont cosignés par 2 à 3 auteurs et 32% sont avec plus de 4 auteurs, ce qui révèle un important degré de collaboration au sein de la communauté scientifique travaillant sur ces questions (cf. l'article d'IBM Research « *Applications of voting theory to information mashups* » avec 9 co-auteurs [14]). Le tableau ci-après présente les principaux auteurs ayant plus de 4 articles à leur actif.

Tableau 1 Les auteurs les plus productifs : une recherche collaborative marquée avec l'industrie et la présence de la Chine en tête

Auteurs	Pub.	Appartenance	Keyword (freq. >= 2)
Meina Song	5	**Chine** : Beijing University of Posts and Telecommunications	Dynamic adaptability, web 2.0, multiple data sources, mobile communication network
Hoyer, V.	5	Suisse : **SAP Research** / University of St. Gallen, Institute for Media and Communications Management,	Enterprise mashup, lightweight composition, web-based resources
Daniel, F.	5	Italie : University of Trento	Web mashups, mashup components, web applications
Maximilien, E.M.	5	USA : **IBM** Almaden Research Center	Web API, domain-specific language
Gomadam, K.	5	USA : Wright State Univ. (travaille avec **IBM** aussi)	SA-REST, Web 2.0, Web API, programmable Web, web sites
Nan Zang	4	USA : Pennsylvania State University	Web-active end user
Sheth, A.P.	4	USA : Wright State Univ. (travaille avec **Accenture**)	SA-REST

- **68% du corpus provient de communications issues de conférences** (224 communications contre 99 publications dans des revues), **majoritairement américaines** (représentant 30% des conférences citées notamment ayant eu lieu en Californie – Los Angeles et San Diego, lieux d'implantation des startups incontournables du domaine comme JackBe ou Kapow Technologies), arrive en second la Chine (12%), et en troisième la France (avec 6%)[14], puis par ordre décroissant l'Australie, l'Autriche, l'Espagne, l'Inde, la Grèce… Pour ce qui est du classement du top des revues, nous trouvons essentiellement des magazines en ligne avec en tête *IEEE Internet Computing* (avec une fréquence de 8), *InformationWEEK* (fréq. de 6), *Searcher* (fréq. de 5) et la revue *Online* (Fréq. de 3), ce qui vient aussi conforter l'idée d'une recherche sur les mashups proche du marché.

[13] Base de données dédiée au STIC avec plus de 11 millions d'articles scientifiques, actes de colloques et rapports techniques indexés (via la banque de données EBSCO de l'INIST), date de l'extraction : avril 2010

[14] Il s'agit essentiellement de communications issues de deux conférences ayant lieu sur Bordeaux : 2009 13th International Conference on Intelligence in Next Generation Networks (ICIN): 'Beyond the Bit Pipes' et 12th International Conference on Intelligent Service Delivery Networks. ICIN 2008

Figure 2a Réseau des mots-clés (champ KW des auteurs) avec les affiliations appartenant à des sociétés

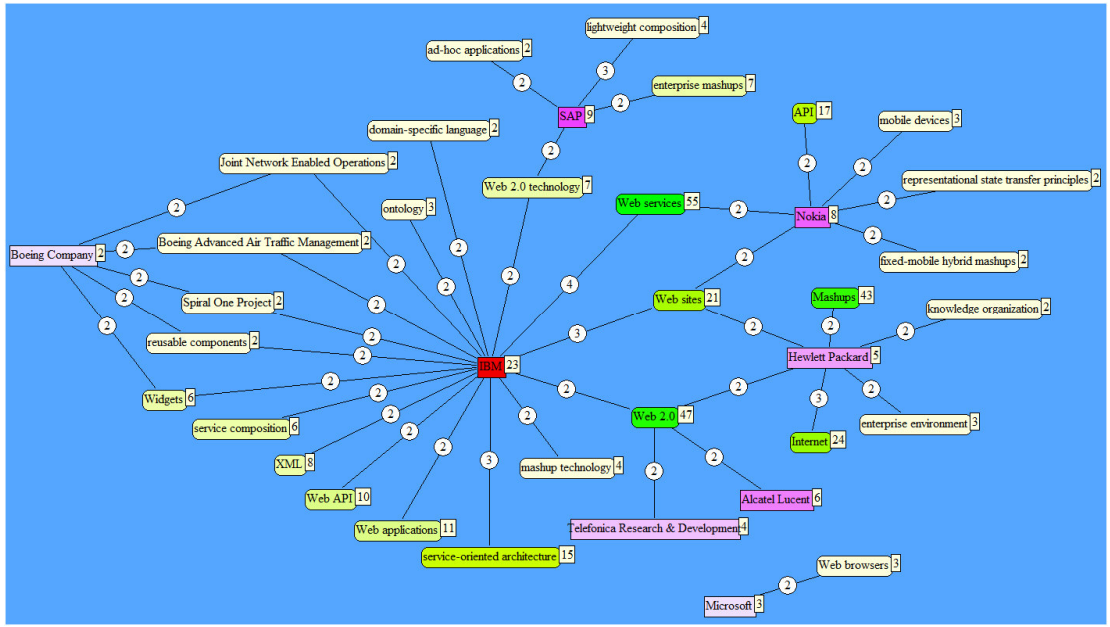


Figure 2b Réseau des articles des sociétés affiliées répartis par année : présence dominante d'IBM R&D sur ces champs de recherche

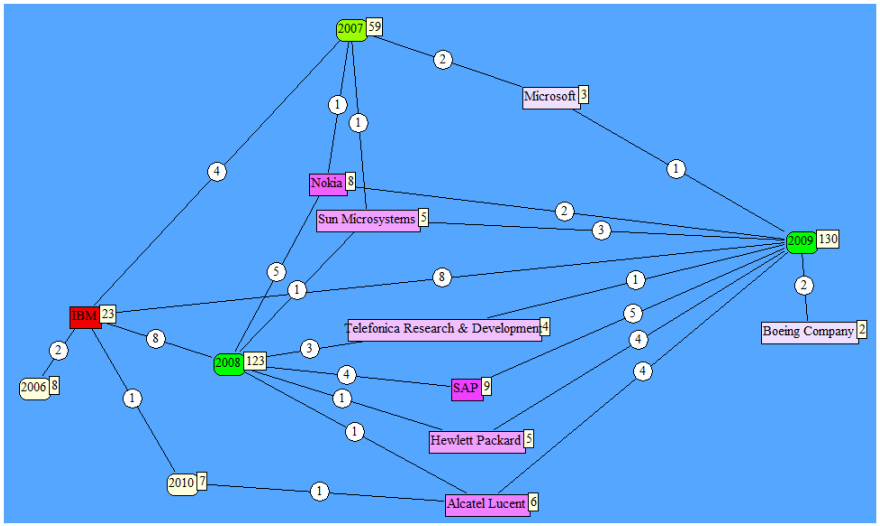


Figure 2 Positionnement des sociétés leaders dans les mashups

- **Plus de 50%** du top 20 des affiliations des auteurs classées par fréquence (soit 65 publications sur 110) **provient de centres de R&D privés** (à savoir, par ordre d'importance, IBM largement en tête avec 23 publications, SAP, Nokia, Alcatel Lucent, Hewlett Packard, Sun Microsystems, Telefonica Research & Développent, Microsoft et Boeing Company). Il est intéressant de souligner que parmi les 8 publications datant de 2006, 50% sont déjà issues de monde de l'industrie, deux proviennent des centres de recherche d'IBM (cf. [15], [16]), l'une de Capgemini [17] et l'une du groupe F5, possédant une grosse communauté d'utilisateurs exploitant les technologies de mashups de la société F5[15] [18]. La Figure 2 représente le réseau de mots-clés co-occurents (freq > 2) de ces sociétés et la répartition de leur production scientifique sur les 5 dernières années ; ce réseau met notamment en évidence une collaboration récente entre IBM et Boeing sur des applications de management de trafic aérien utilisant les mashups [19]. Nous voyons également sur ces graphiques les orientations des recherches effectuées par SAP et HP autour de déploiement de mashups dans l'entreprise, dédiés au BPM et au marché de l'entreprise 2.0 en général [20][16].
- **Une recherche qui intéresse de plus en plus l'Asie** : S'il est indiscutable que le cœur de la recherche sur les mashups est positionné aux Etats-Unis, les pays asiatiques cherchent aussi à se placer dans la course ; la Chine arrive en second après les Etats-Unis en tant que pays d'accueil pour l'organisation de conférences de notre corpus dont la moitié a eu lieu en 2009, d'autres pays comme l'Inde ou la Corée du Sud commencent aussi à occuper ce terrain[17].

[15] Voir http://devcentral.f5.com
[16] Voir Annexe 7.1 pour visualiser les nuages de mots-clés des articles des principales sociétés extraites du corpus
[17] Voir Annexe 7.2 pour une vue de la liste des conférences sur 2009/2010 obtenue à partir de notre corpus

Nous avons cherché à cartographier le corpus sous différents angles afin de mettre en évidence les champs d'application et les centres d'intérêts de la communauté scientifique sur la question des mashups. Dans l'esprit de [21] cherchant à caractériser les centres d'intérêt de la recherche sur les réseaux sociaux au moyen d'analyses scientométriques, nous avons construit un premier réseau de cooccurrences de mots-clés issus du champ « Key Phrase Headings » et bien qu'il n'y ait pas de clusters de mots-clés très significatifs, nous retrouvons sans grand surprise une forte corrélation entre les termes *Mashup*, *Web Services* et *Web 2.0*, autour de termes décrivant l'essentiel des technologies caractérisant le paradigme du web 2.0[18] (comme *Social Neworking*, *Semantic Web*, *Wikis*, *Blogs*, *Service-oriented Architecture*, *API*, *AJAX*, *Widgets*,…). Pour approfondir, nous avons croisé les codes de classification d'INSPEC correspondant à des champs d'application définis (comme la santé, la finance, l'administration publique), avec les mots-clés, en vu de faire émerger des tendances plus originales que celles précédemment détectées.

Figure 3a Cartographie des cooccurences de mots-clés (champ KW) hors fréq. =1

Figure 3b Cartographie du croisement des domaines d'applications issus des codes de classification de l'index d'INSPEC (champ CC) avec les mots-clés (champ KW) hors fréq.=1

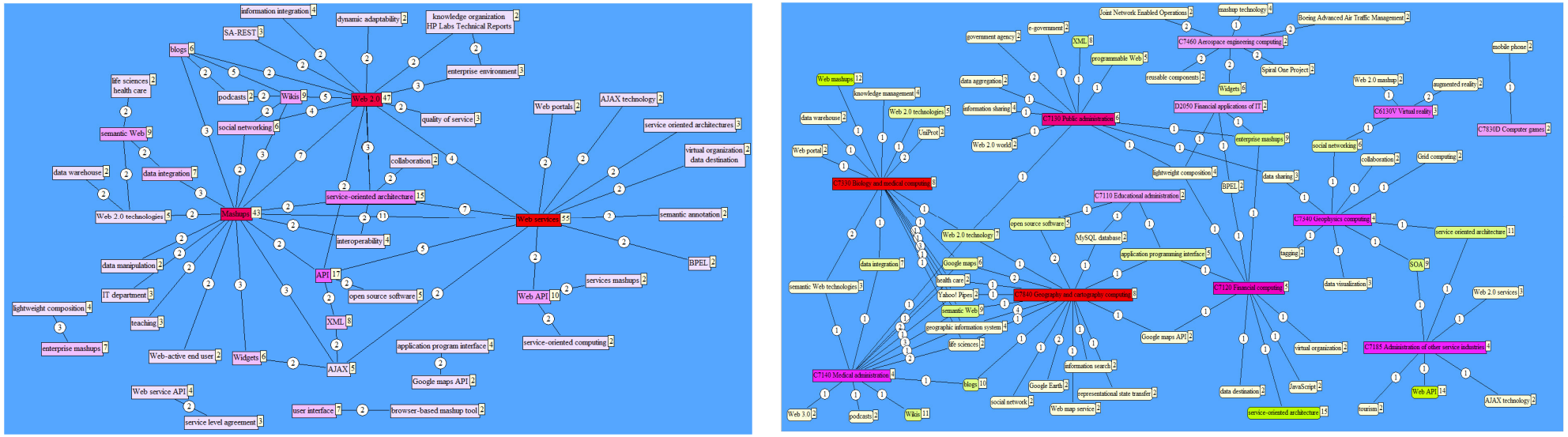


Figure 3 Cartographie des domaines d'applications d'INSPEC et mots-clés des auteurs : les mashups d'entreprise cristallisent les intérêts de recherche

Deux éléments forts se distinguent, d'une part la prédominance des mashups de géolocalisation avec des applications notamment dans les domaines du tourisme et l'industrie du voyage[19], de la santé (avec les mots-clés *health care education*, *eHealth*) ou de la finance et d'autre part l'utilisation importante de mashups appliqués au monde de l'entreprise (ie *business mashup*) déployés dans le monde de la finance, la gestion d'entreprise, l'administration publique ou la santé (cf. les termes *BPEL*, *ligthweight composition*, *SOA*, *data/information sharing*). Notons que nous retrouvons également la présence importante de ces deux domaines dans l'analyse du corpus de ProgrammableWeb. Nous détaillons dans les paragraphes suivants des exemples de combinaisons de mashups, extraits des tendances mis en lumière lors de cette analyse.

[18] Cf. le blog de référence sur le web 2.0 de Dion Hinchcliffe : http://web2.socialcomputingjournal.com/

[19] Cf. l'application originale de système de recommandation intelligent basé sur des web-services sémantiques et appliqués au tourisme [50)

3.2 Premiers résultats : tendances fortes sur les « Next-Generation Mashup »

Dans le but de dégager des tendances émergentes, notre analyse s'est d'abord portée sur les derniers articles parus (soit 7 publications datant de 2010 lors de notre extraction).
Même si il est difficile de tirer des conclusions générales, il est tout de même intéressant de relever ces quelques éléments :

- des applications de mashups connectées à la veille avec deux articles signés d'universités indonésiennes proposant des solutions d'extraction de données web par composition de mashups, dont l'une est complètement dédiée au cycle de veille par la combinaison de mashups d'extraction de données brutes, de filtrage et d'affinage, d'homogénéisation des sources et des données et finalement de visualisation des données agrégées [22] et l'autre est appliquée à la scientométrie et l'analyse de citations avec la solution Robomaker de la société OpenKapow [23]. La présence de l'Indonésie vient étayer notre précédente remarque sur la prise de position manifeste des pays asiatiques dans la recherche sur les mashups.
- la recherche de solutions pour sécuriser la composition de mashups avec deux articles provenant d'une part d'IBM USA sur l'utilisation d'annotation sémantique pour organiser la composition de web services [24] et d'autre part d'Alcatel Lucent France sur un modèle de réputation pour sécuriser l'accès et l'utilisation de mashups [25][20].
- enfin les autres articles concernent une application touristique de géolocalisation de données et un article du MIT sur une application de mashups pour la combinaison de données provenant des réseaux sociaux, de télévision sociale et de découvertes de données collectives [26].

Notre attention s'est ensuite portée sur le terme explicite « next-generation mashup » (avec une fréquence de 8 dans notre corpus), trouvé notamment plusieurs fois dans les articles provenant d'IBM (cf. Annexe 7.1 Tagcloud des sociétés publiantes). Nous avons ainsi pu dégager quelques tendances d'avenir autour des mashups :

- un intérêt marqué comme axe de recherche sur les limites reconnues du déploiement des mashups dans l'entreprise comme alternative de remplacement des lourdes plateformes applicatives type corporate groupware ou de portail centralisé d'entreprises. Comme nous l'avons vu précédemment, même si les études marketing des instituts comme Forrester ou Gatner prévoient l'envolée imminente du marché des mashups, leur implémentation est encore freinée par des problèmes de scalabilité et d'interopérabilité des données, de sécurité et de confiance[21] dans le mixage et la réutilisation des mashups créés, d'open source et de données ouvertes versus propriétaires, de qualité de services de bout en bout dans la chaine d'orchestration des mashups mixant des données hétérogènes appartenant à des entités différentes[22]. Il est d'ailleurs intéressant de souligner la présence constante d'IBM (des centres de new-York, d'Allemagne et de Taiwan) tant sur le plan technologique que juridico-économique autour des questions de « *next generation enterprise portals* » et l'orchestration sécurisée de web services (à noter le néologisme « *SOA-lization of legacy and security applications* » dans [27] et les travaux approfondis sur les limites juridico-légales qu'implique la réutilisation de mashups d'entreprises « *accountable enterprise mashup services* » [28]).
- nous retrouvons Ericsson en partenariat avec Telefonica (Espagne) sur une tendance qui mixe la question de l'orchestration de mashups avec les technologies mobiles : ces derniers travaillent sur une plateforme de combinaison self-services de mashups mobiles centrés utilisateur [29]. Citons aussi les travaux très novateurs de l'université de Beijing sur la reconfiguration de mashups mobiles par la mise en place d'un framework de gestion de qualité de service complètement contextualisé et géo-centré sur l'utilisateur en état de mobilité (ie les termes *feedback-enhanced QoS model*, *QoS-oriented management framework* ou *reconfigurable mobile mashup services*), ou encore NTT Corp Japan qui travaille sur une plateforme de réalité virtuelle mixant avatar, fonctionnalités sociales du web 2.0, messagerie instantanée [30]. Pour finir sur cette tendance marquée de l'évolution des mashups vers les technologies mobiles, citons une recherche d'avant-garde de chez Nokia sur une plateforme de mashups mixant des données de capteurs provenant d'utilisateurs où le mobile est à la fois source de données capteurs et moyen d'accès et d'interaction avec la plateforme de services [31].

---

[20] Il n'est d'ailleurs pas anodin de souligner ici le rachat de Programmableweb par Alcatel Lucent fin juin 2010 : http://blog.programmableweb.com/2010/06/29/programmableweb-joins-alcatel-lucent/
[21] Retrouvé sous l'acronyme trois A « Authentication, Authorization and Accounting » (Stecca, Maresca, & Baglietto, 2009)
[22] ie des termes comme *Quality of Services – QoS* ou *Service level Agreement – SLA* [45] ayant des fréquences supérieures à 4 dans le corpus

Concernant la relation explicite entre les mashups et la veille, nous n'avons pas trouvé de termes significatifs dans les mots-clés du champ KW, mis à part *competitive intelligence*, apparaissant qu'une fois dans un article de Microsoft sur des solutions de Business Intelligence orientées web services et web 2.0 mises en place dans l'industrie pétrochimique [32]. Cependant, même si en terme de fréquence, *learning* n'apparait que 6 fois dans nos articles, nous avons également repéré des travaux effectués par des universités allemandes et autrichiennes mais aussi chinoises autour des Environnements d'Apprentissage Personnel (ie *Personal Learning Environment - PLE*) : les mashups constituent une alternative enrichissante pour palier aux plateformes de management de connaissances classiques (ie *Learning Management System - LMS*), jugées rigides, top-down et ne répondant pas aux besoins fluctuants des utilisateurs. Nous pensons que ces applications sont fortement corrélées aux applications pour la veille, plusieurs travaux rapprochent en effet le processus de veille avec les questions d'apprentissage, de capitalisation des connaissances ou plus largement de création de sens (ie *sense making* [33] ou *sense meaning process* [34]).
Plusieurs prototypes de combinaison de mashups sont actuellement testés, comme PLEF [35], un framework développé en collaboration entre l'université de Aachen (Allemagne) et l'Open University de Heerlen (Pays-bas) proposant une collection de web services pour aider les apprenants à construire leur propre réseau de connaissances (ie Personal Knowledge Networks - PKN), ou MUPPLE - *Mash-UP Personal Learning Environment* [36] de l'université de Vienne (Autriche) basé sur des mashups d'interaction sociales et de partage de meilleures pratiques (ie *pattern-based best practice sharing*) ou encore Jampots [37] issu d'une collaboration entre deux universités chinoise et américaine, qui propose un écosystème d'apprentissage basé sur l'interopérabilité des services, leur auto-régulation, leur réutilisation, leur facilité d'utilisation, en étant orienté utilisateur. Cependant, l'un des assemblages de mashups les plus prometteurs provient d'une recherche sino-autrichienne, le projet ROLE - *Responsive Open Learning Environments* [38] qui vise à utiliser les technologies mobiles dans le déploiement d'environnements d'apprentissage basés sur les mashups pour qu'ils soient totalement flexibles, auto-adaptatifs, en mode self-service centré utilisateur et surtout ouverts au plus grand nombre, l'université de Shanghai Jiao Tong University doit être l'un des premiers terrains d'expérimentation de ce nouveau type d'Environnement d'Apprentissage Personnel Mobile.

Dans ce chapitre consacré à l'analyse bibliométrique d'un corpus issu d'INSPEC sur le terme exact « mashup », nous avons pu mettre en évidence l'importance de la recherche sur les mashups d'entreprise, déployés en écosystème de web-services[23] pour répondre, entre autres, à des besoins de Business Intelligence, et le positionnement marqué des leaders du marché sur ces terrains comme IBM, SAP, HP ou Nokia [39]. Le parcours fouillé des mots-clés traités et isolés des titres des articles comme des champs spécifiques de la base INSPEC nous a permis aussi de dégager certaines tendances originales de combinaison de mashups (ie le mix des fonctionnalités autour de *knowledge management* / *elearning* / technologies mobiles / données de géolocalisation, de réseaux sociaux/ analyse de capteurs, de sentiments…). Nous avons aussi pu repérer des projets de recherche qu'il est important de prendre en considération au regard du projet ISICIL.
Cependant, nous attirons ici l'attention sur le fait que notre approche n'a pas vocation à être une étude exhaustive sur le cadre de recherche général dans lequel s'ancrent les mashups (comme sur les questions d'open innovation centrée utilisateur, d'innovation dite démocratique ou recombinée cf. les travaux de Von Hippel sur la caractérisation des premières plateformes de création de mashups, antérieurement connues sous le terme « boite à outils agile et sociale » - *User Toolkits* [40]).
Il s'agit d'une première analyse exploratoire dans un objectif précis : celui d'effectuer un recensement et une sélection de mashups et APIs utiles pour des besoins de veille. Dans ce but, nous avons poursuivi notre analyse du marché des mashups dans une démarche complémentaire au positionnement scientifique en s'appuyant sur les métriques d'analyse des réseaux sociaux en vue faire émerger des combinaisons effectives, voire singulières de mashups et d'APIs pour outiller le veilleur.

[23] Tel que l'entend West, composé de fournisseurs de technologies et de composants, de fournisseurs de plateformes intégratrices, d'agents proposant des services aux fonctionnalités spécifiques et en bout de chaîne les utilisateurs [52].

4 Le marché des mashups : analyse du répertoire en ligne ProgrammableWeb

4.1 Méthodologie adoptée : l'analyse des réseaux sociaux pour faire émerger des combinaisons singulières

Nous avons choisi d'analyser les données du site ProgrammableWeb.com d'une part pour la richesse et l'actualité de son répertoire avec plus de 2000 APIs et 5000 mashups répertoriés (soit 3 nouveautés par jour en moyenne), et d'autre part pour l'accès aux données brutes du répertoire au format xml (REST/JSON) via son interface api.programmableweb.com, facilitant l'interrogation de données suivant des critères de sélection par tags, par apis, par date, par popularité, etc…

Tableau 2 Résumé des résultats de l'analyse bibliométrique des fiches de mashups issues de ProgrammableWeb

Résultats d'analyse du corpus des 2016 fiches de Mashups							
Sélection sur toute la période 2005 – 2010				Sélection sur 2010 exclusivement			
Top des APIs les plus combinées		Top des Tags		Top des APIs 2010		Top des Tags 2010	
APIS	Fréq.	Tags	Fréq.	APIS	Fréq.	Mashups	Fréq.
Google Maps	428	search	700	Twitter	17	Search	17
Flickr	291	social	517	Flickr	10	social	13
Twitter	274	mapping	479	Google Maps	9	visualization	11
YouTube	213	photo	331	Google Chart	8	photo	10
del.icio.us	186	visualization	269	Amazon eCommerce	8	charts	8
Amazon eCommerce	178	video	258	Wikipedia	6	mapping	7
Facebook	143	shopping	237	Twilio	6	shopping	7
Yahoo Search	139	bookmarks	216	Facebook	6	microblogging	7
Google Search	135	news	203	YouTube	5	**telephony**	6
Google Ajax Search	106	reference	200	LinkedIn	4	video	5
Last.fm	87	rss	157	Google Visualization	4	semantic	5
eBay	86	microblogging	149	eBay	3	enterprise	5
Technorati	78	music	138	del.icio.us	3	bookmarks	4
Digg	77	messaging	132	Yelp	3	wikipedia	4
Google Chart	65	enterprise	109	Yahoo BOSS	3	**socialgraph**	4
Wikipedia	52	charts	104	Shopping.com	3	auction	3
Yahoo BOSS	49	widgets	101	Meaningtool	3	alerts	3
Microsoft Virtual Earth	44	wiki	97	Google Search	3	Voice	3
Yahoo Image Search	44	semantic	96	DBpedia	3	statistics	3

L'extraction complète de la base de données des mashups de Programmableweb via son API étant fastidieuse (les résultats sont fixés à 100 fiches par page maximum), notre méthodologie a été de sélectionner, à partir du Tagcloud des mashups enregistrés, quelques 150 tags significatifs en nous basant sur nos connaissances des fonctionnalités du processus de veille présentée au paragraphe 2.3 (ie par le choix de tags comme *collaboration*, *community*, *search*, *brainstorm*, *mindmap*, *clustering*, *dashboard*, *microbloging*, *bookmark*, *semantic*, etc…), nous avons ensuite interrogé l'API pour chaque tag afin de récupérer 2016 fiches de mashups. Bien que le format xml d'une fiche propose de multiples champs potentiellement exploitables en datamining comme le champ auteur, le nombre de commentaires, le nombre de téléchargements, la notation des utilisateurs, etc, nous nous sommes focalisés sur les champs Tags et APIs des 2016 fiches de mashups (soit 1549 mashups après dédoublonnage). Nous avons pu extraire 649 tags (avec plus de 50% des fiches de mashup taguées de 3 ou 4 tags, plus de 25% avec 5 tags et plus, et seulement 2% avec un tag, ce qui est bien au dessus des comportements de tagging remarqués dans les services de web 2.0). Le tableau 2 présente les classements des APIs et des tags les plus utilisés, d'une part sur toute la période 2005-2010 et spécifiquement sur l'année 2010 pour faire ressortir les tendances récentes. Les mashups sont reliés à 457 APIs dont en tête les APIS les plus populaires du répertoire Programmableweb comme Google Maps, Flickr, Twitter et

YouTube avec en plus dans les APIs leaders Facebook et Delicious pour les réseaux sociaux et Google Search et Yahoo! Search pour la recherche d'informations, puisque nous avons sélectionné des mashups dont les tags sont de près ou de loin liés aux activités de veille. Il est à noter entre les deux périodes une stabilité dans la présence de certains tags au top comme *Search*, *social*, *mapping*, *visualization* ou *microbloging* (entretenu par le phénomène Twitter qui fait partie des cinq premières APIS les plus populaires de répertoire Programmableweb), avec deux tags qui se distinguent en 2010, l'apparition de *Telephony* (et non *mobile*, qui dans notre corpus est cité 56 fois par les mashups sélectionnés) et *socialgraph* (pour représenter visuellement les réseaux sociaux).
A ce stade de l'analyse, nous repérons tant au niveau des mashups que des APIs, qu'il est possible de rapprocher ces deux groupes autour de tags significatifs pour savoir si il existe effectivement des combinaisons incontournables, singulières ou innovantes propres à outiller le veilleur. A partir du croisement des champs xml des fiches, nous avons donc extrait deux principales matrices « label des Mashups / label des APIs » et « label des Mashups / Tags associés ».

4.1.1 Premières révélations sur la structure globale du réseau : les mashups incontournables pour la veille

Dans le but d'approfondir cette première analyse bibliométrique mais aussi inspiré par les travaux exploratoires de Yu [41], nous avons choisi d'appliquer les métriques d'analyse des réseaux sociaux (ie *Social Network Analysis - SNA*) sur notre corpus, notamment pour découvrir s'il n'existait pas des schémas types d'interaction entre les mashups de notre sélection et les APIs auxquelles ils sont reliés. Nous avons donc souhaité comprendre : comment les mashups étaient reliés aux tags qui les décrivent, quelles sont les APIs qui jouent un rôle critique dans l'écosystème de Programmebleweb, quels tags ont suscité le plus d'intérêt, ou encore comment les APIs sont connectées entre elles. Pour ce faire, nous avons sélectionné certaines métriques comme la centralité de degré (zoom sur le nombre de connexion par nœud du réseau), la centralité de prestige (nous choisirons l'indice d'eigenvector pour repérer les nœuds connectés aux nœuds les plus prestigieux en terme de connexion) et l'analyse des liens avec PageRank (autre mesure de la popularité d'un nœud dans le réseau) que nous propose l'outil NodeXL [42].
La série de figures suivantes (cf. Figure 4) présente le réseau de Mashups reliés aux APIs qu'ils combinent des 1549 fiches sélectionnées, en utilisant comme représentation graphique le layout *Harel-Koren* de représentation multi-couche des nœuds (pour affiner le graphe et regrouper les nœuds proches par similarité). Nous avons réduit le bruit de ces 1979 nœuds générant 3673 liens en enlevant successivement les nœuds les moins denses en liens, visibles en bas du graphe, pour finalement faire ressortir au cœur du réseau le nœud dense de mashups et d'APIs les plus populaires et dont les combinaisons sont les plus courantes.

A titre de comparaison, nous nous sommes référés aux travaux exploratoires de [41] et son analyse longitudinale entre 2005 et 2007 sur l'évolution de la base de mashups de Programmableweb au moyen d'analyses de réseau sociaux. Chez Yu, la représentation graphique du réseau obtenu des relations entre Mashups et APIs montre une structuration du réseau en trois tiers, au centre, les APIs et mashups utilisant GoogleMaps (utilisée par près de 40% des mashups actuellement), autour du centre se trouvent les APIs populaires dédiées au management et la recherche de données sociales et plutôt de type plateforme de services (ex Facebook, Twitter, Flickr, Amazon) et finalement en périphérie les APIs moins populaires, fournisseurs de données brutes (vidéo, photo, musique) pour des applications type feeds, ecommerce, musique… Vu que notre corpus sélectionné ne représente pas la totalité des mashups de ProgrammableWeb, nous pouvons voir que l'API GoogleMaps bien qu'étant la plus populaire en nombre de liens n'est pas tout à fait centrale et qu'il figure aussi au centre de notre corpus les APIs de recherche, d'indexation (via des tags), ou de traitement de données comme Del.Icio.Us, Google Search, Yahoo Search, Digg, Technorati ou Google Chart (outre la présence aussi dans le cercle des APIs populaires Twitter, Flickr, Youtube, Amazon Ecommerce ou Facebook).

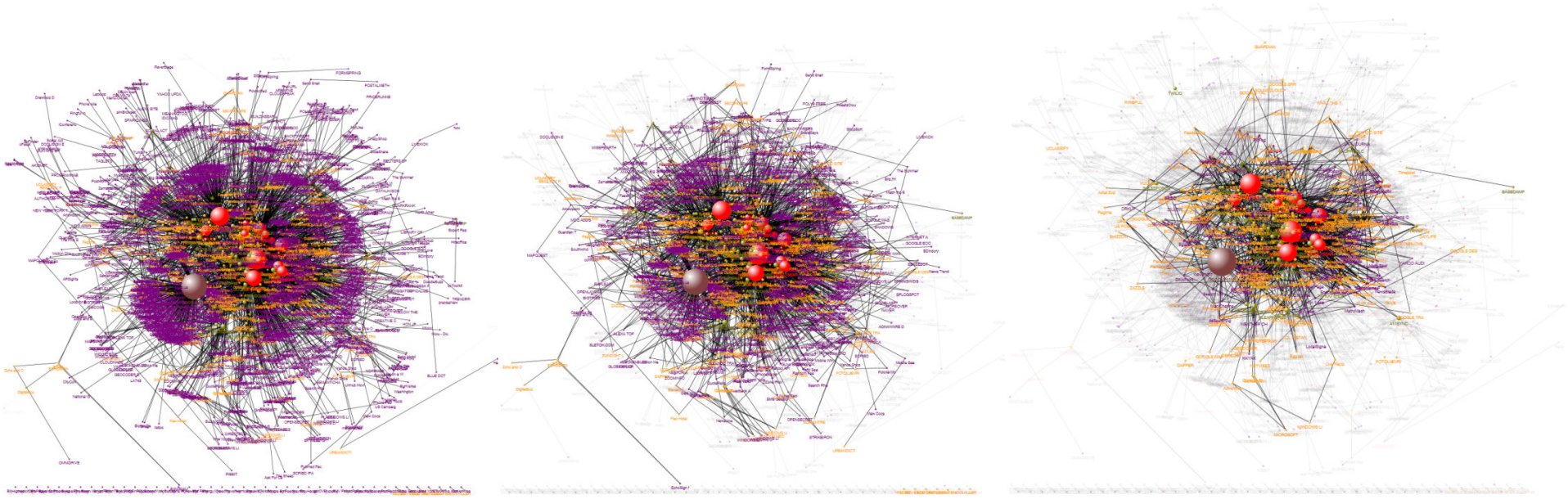


Figure 4 Représentation de la structure globale du réseau Mashups-APIs : Les APIS sont représentées par une sphère, les mashups par un losange, la taille est fonction du calcul de centralité de degré, la couleur correspond au calcul de popularité des nœuds dans le réseau avec la métrique du PageRank (mis à part GoogleMaps l'API la plus populaire en marron, le nombre de liens décline du rouge, olive, orange et violet pour les moins populaires)

Cependant, pour nous aider à cibler plus précisément ces relations entre mashups et APIs notamment en vue d'outiller le processus de veille, nous nous sommes tourné vers la matrice Mashups / Tags, afin de discriminer les mashups en fonction des tags qui se rapprochent le plus des activités de veille. Une première analyse du réseau obtenu entre les 1549 mashups et les 649 tags donnant lieu à un réseau trop dense (avec plus de 5570 liens), nous avons effectué une sélection de 60 tags significatifs au regard de notre objet d'étude et nous avons lancé le calcul de détection de cluster[24] sur cette nouvelle matrice Mashups / sélection de Tags. Nous avons pu mettre en évidence 13 familles de mashups regroupées par tags associés et présentées au tableau 3. Certaines associations ne nous étonnent pas comme celles caractérisant typiquement les fonctionnalités des applications web 2.0 (ie clusters « Bookmark social, tagging social », « Fonctionnalités autour du blog » ou « Réseaux sociaux et web sémantique ») ou les fonctionnalités ciblées entreprise 2.0 (ie clusters « Business intelligence », « Visualisations graphiques » ou « Analyses, datamining »).

D'autres clusters sont plus étonnants, comme celui reliant les tags *analysis*, *microbloging*, *sentiment* et *trends* où l'on trouve des mashups exploitant essentiellement l'API de microbloging Twitter dans le but d'extraire toutes sortes de tendances (cf. les mashups Linksensor[25], Twitterthoughts[26], Twitgraph[27] ou Google Hot Trends + Twitter + Google Maps[28]). Autre ou le cluster reliant les tags *finance, dictonary, answers, mobile*, où nous retrouvons la tendance de combinaisons de mashups de

---

[24] Sélection dans NodeXL de la méthode de clustering Wakita-Tsurumi, donnant de meilleurs résultats que celle par défaut de Clauset-Newman-Moore sur notre corpus Mashups/tags

[25] *LinkSensor* ***semantically analyzes blog posts or articles*** *on-the-fly and* ***highlights key concepts in the text***

[26] *TwitterThoughts visualizes the tweet stream by creating a* ***tweet motion chart****, the* ***twitter world map*** *and tweet thoughts break out within a time period*

[27] *Twitgraph tracks micro-trends on twitter and brand user satisfaction. See metrics visualized using Google Visualization API.*

[28] *Google Maps Mashup of Twitter Searches which contain* ***keyword*** *from Google Hot* ***Trends***

données avec des applications mobiles comme le mashup Ixcitable for Jigsaw d'accès via l'Iphone à l'annuaire d'entreprise et de cartes de visite Jigsaw ou Dbpedia Mobile ou Mobile EVDB Events Search de recherche via le mobile dans des répertoires DBpedia et d'événements EVDB. Bien que Yu l'aborde seulement dans la conclusion de son étude exploratoire, il démontre que, malgré le succès des quelques 20% d'APIS les plus populaires qui génèrent 95% des créations de mashups, ce pourrait être du côté des APIs appartenant à la périphérie, que les combinaisons Mashups/APIs seraient innovantes et créatrices de nouvelles fonctionnalités. Nous présentons en annexe 7.4 la cartographie du réseau Mashups/sélection de Tags formée par ces clusters, mettant graphiquement en évidence pour chaque groupe identifié les mashups centraux considérés comme les plus populaires, mais aussi les mashups périphériques, auprès desquels nous pourrions trouver des combinaisons singulières, donc par extrapolation innovantes. Nous avons ainsi pu trouver les mashups comme Nobosh[29] reliant les tags *business* (à forte popularité), *community*, et *aggregator*, Simple Del.icio.us Spy[30] reliant *slideshow* à *delicious*, Mytag[31] reliant *tagging* et *personalization* ou encore FOAFster[32] reliant *foaf* à *socialgraph*.

Tableau 3 Clusters thématiques de combinaisons de mashups hors nœuds de degré 1 (Cf. annexe 7.4 pour sa représentation graphique)

Titre général des clusters	**Couleur**	**Nbr. Mashups**	**Tags associés**
Fonctionnalités autour du blog	Losange violet	46	*blog, rss, comments*
Réseaux sociaux et web sémantique	Sphère rouge	35	*people, SemanticWeb, semantic, foaf, research, socialgraph, personalization, related*
Visualisations graphiques	Sphère orange	32	*visualization, graphics, slideshow, comparison*
Analyses, datamining	Sphère magenta	30	*charts, metrics, analytics, crouwdsourcing, datamining*
Bookmark social, tagging social	Sphère olive	28	*delicious, ratings, tag, bookmarks, tagging, aggregator, community*
Business intelligence	Sphère bleu	27	*directory, enterprise, documents, Serena, marketing, business*
Fonctionnalités mobiles de CRM, finance	Triangle bleu	25	*finance, dictonary, answers, mobile*
Analyses, prédiction, microblogging	sphère anthracite	25	*analysis, microbloging, sentiment, trends*
Cartographie collaborative, traduction, wiki	Losange marron	22	*maps, translation, wikipedia, wiki*
Recommendation d'événements	Losange marron	18	*events, recommandations*
Systèmes d'alerte et de réputation	Losange magenta	16	*alerts, reputation*
Elearning, recherche 2.0	Sphère violette	14	*collaboration, education, university, database*
Réseauting	Losange rouge	8	*Networking*

Ces premières représentations globales des réseaux Mashups/APIs et Mashups/Tags nous ont permis de mettre en lumière l'existence de mashups dont les fonctionnalités répondent pleinement à des besoins de veille (recherche d'experts via les réseaux sociaux, détection de news, visualisation graphique de tendances, etc…). Nous présentons la suite de notre démarche itérative dans le but final de dégager des combinaisons intéressantes de mashups et d'APIs pour outiller le veilleur.

[29] *Nobosh finds **news stories for everything from the latest current business news**, tech news, health news, advertising news, and corporate blogs powered by Nobosh algorithms*

[30] *Simple Del.icio.us Spy is based on the Del.icio.us API and jQuery library. The box **monitors recent bookmarks saved by users** using a cute slide show effect to show updates*

[31] *MyTag A personalized cross-media **folksonomy search engine***

[32] *FOAFster A prototype visualizer for Friend Of A Friend relationship objects. It mashes up MyBlogLog FOAF objects, and **adds favorites** from del.icio.us, Digg, Last.fm, LinkedIn, StumbleUpon, Upcoming, Tumbler, Twitter, and YouTube if it can find them*

4.1.2 Repérage d'éléments singuliers au sein des réseaux

Le fil conducteur de notre démarche a été d'appliquer pour partie la trame d'analyse des réseaux sociaux développée par [43] : il s'agit de partir d'une approche globale du réseau obtenu afin de dégager si il ne se détache pas des sous-groupes distincts dans le réseau global, pour aller vers une analyse plus micro, en se concentrant sur les éléments singuliers au sein du réseau justement mis en évidence par les métriques citées précédemment.

Nous avons donc cherché à mettre en évidence des nœuds singuliers de nos réseaux en utilisant les métriques de SNA calculant la popularité pour effectuer un zoom sur les mashups les plus représentatifs de notre corpus. La figure 5a présente une vue globale de la matrice « labels des mashups / tags » avec le layout d'Harel-Koren, intéressant pour mettre de l'ordre dans les nœuds du réseau en les regroupant par similarité, nous retrouvons regroupés autour des tags significatifs nos familles de mashups.

Figure 5a Réseau Mashups/sélection de tags représenté avec le layout Harel-Koren (les nœuds sont assemblés par similarité)

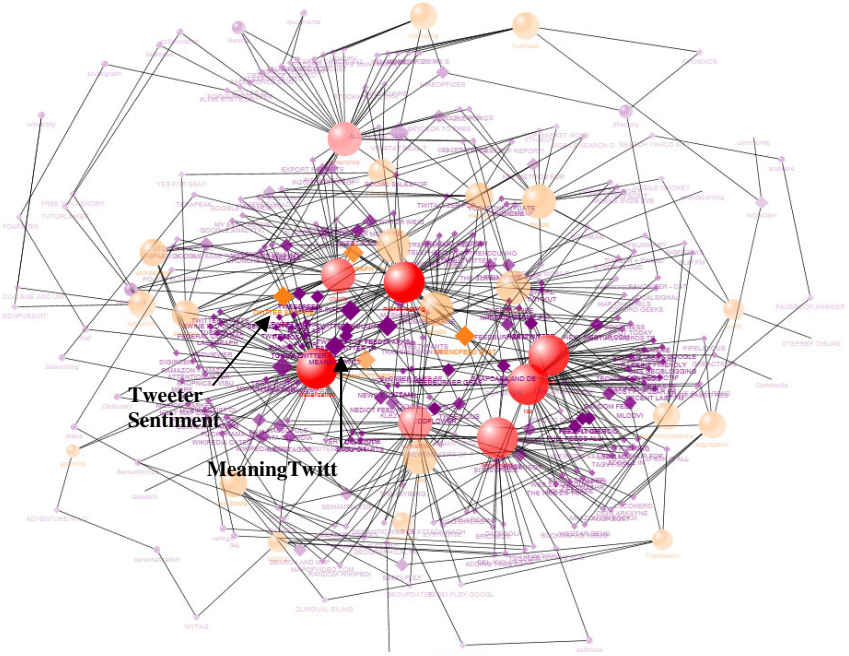


Figure 5b Mise en relief du réseau de mashups autour du tag Enterprise (représentation des liens de niveau 1,5)

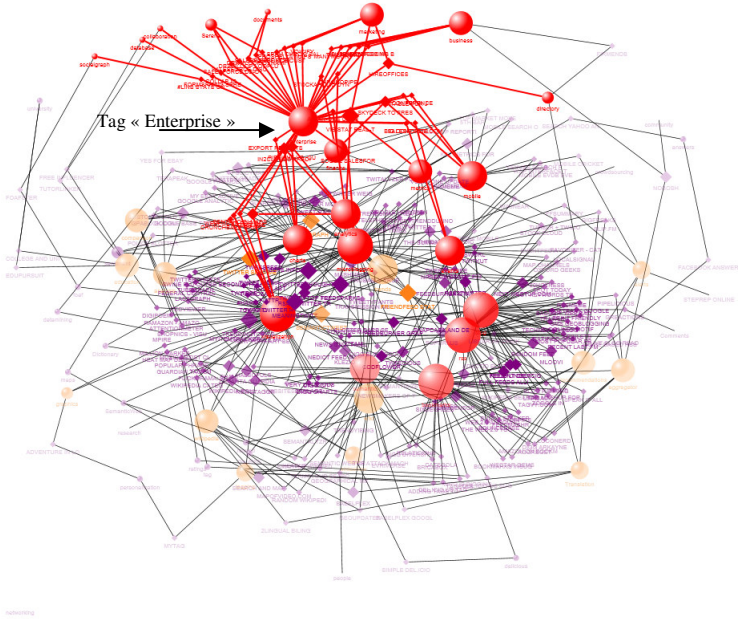


Figure 5 Réseau représentant les mashups (losange) reliés aux tags (sphère) les plus populaires (freq > 2). Pour plus de clarté les nœuds de degré 1 ont été éliminés de l'analyse et les nœuds de degré 2 et 3 sont réduits en opacité de 5%. La taille est associée à la centralité de degré (soit les plus gros sont les plus fournis en liens), l'opacité est reliée à la centralité de prestige (ie Eigenvector centrality *: moins le prestige est grand, moins les nœuds sont visibles), la couleur est corrélée au coefficient Pagerank de calcul de notoriété dans le réseau (rouge pour les plus populaires, et orange pour l'intermédiaire et violet pour un pagerank sous la moyenne)*

Dans la représentation globale du réseau figure 5a, nous retrouvons bien au centre les sphères rouges des tags les plus représentatifs et populaires de notre corpus comme *microbloging*, *visualization*, *blog*, ou encore les tags *rss*, *charts*, *enterprise* ou *semantic*, toujours centraux mais moins prestigieux (c'est-à-dire qu'ils sont reliés à des mashups combinant des tags plus rares donc moins populaires, au contraire de mashups comme MeaningTwitt[33] qui est relié aux tags *visualization*, *semantic*, *microblogging* et *charts* ayant tous une forte centralité de degré) ou Tweeter Sentiment[34] (reliant *microblogging*, *visualization*, *charts* et *sentiment*). Les losanges de grandes tailles et bien visibles se trouvant en position plus ou moins centrales autour des sphères rouges et oranges de grande taille représentent quant à eux les mashups à ne pas manquer, c'est-à-dire reliés aux tags les populaires donc les plus utilisés au sein de ProgrammableWeb, comme Klezio[35], MashMeUp[36], LinkSensor (déjà repéré), Sisense[37], NewsTaggr[38], ou encore Feedz[39].
Cette visualisation nous permet de détecter les mashups relativement connus sur Internet et dont les fonctionnalités peuvent aider le veilleur dans son travail d'analyse. Pour aller plus loin dans le repérage de combinaison moins évidentes mais tout aussi intéressantes de mashups pour la veille, nous avons porté notre attention sur les mashups associés aux sphères oranges représentant les tags populaires en nombre de mashups qui les citent, mais dont les mashups eux-mêmes sont reliés à des tags peu utilisés (d'où leur visibilité amoindrie par l'indice d'Eigenvector). Pour exemple, nous avons représenté dans la figure 5b le réseau de liens directs et indirects entre le tag « enterprise », les mashups qui lui sont associés et les autres tags également impliqués. Nous pouvons ainsi mettre en évidence que les mashups tagués avec « enterprise » sont aussi tagués avec les tags *analytics*, *business*, *charts*, *collaboration*, *database*, *directory*, *documents*, *enterprise*, *events*, *finance*, *marketing*, *metrics*, *microblogging*, *mobile*, *socialgraph*, *visualization*, et *Serena* (notons la seule présence de Serena, acteur important sur le marché des mashups animant autour de sa plateforme de création de mashups dédiée à la BI, une communauté de créateurs et d'utilisateurs de mashups, un des rares avec Jackbe à proposer cela). Il est intéressant de voir que c'est à l'intersection de tags plus populaires avec ceux moins populaires que nous trouvons notamment des mashups de Business Intelligence comme Webfocus For Google Maps[40] (entre les tags *enterprise* et *visualization*), d'analyse et de datamining comme Visistat Real-Time Web Analytics[41] (entre les tags *marketing*, *metrics*, *analytics* et *enterprise*) ou Stockaholic Dynamic Charts[42] (entre les tags *enterprise*, *metrics* et *finance*).
Nous avons pu, avec cette démarche, cibler les mashups parmi les 1549 fiches, qui nous semblaient utiles pour la veille en les regroupant autour des tags significativement proches des besoins de veille.

4.2 Nos résultats de combinaisons d'APIs et de mashups pour la veille

Finalement, nous avons systématisé l'approche précédente, en confectionnant une nouvelle matrice Mashups/APIs constituée des mashups liés aux tags représentés par les sphères rouges et oranges de la figure 5a, soit autour des tags *enterprise*, *semantic*, *business*, *analysis*, *trends*, *metrics*, *recommandation* et *analytics*. Notre réseau se réduit à l'analyse de 639 liens entre une sélection de 196 APIs et 154 Mashups, affinées grâce à nos analyses précédentes et dont nous sommes sûrs que les combinaisons peuvent être appliquées aux besoins du veilleur.

---

[33] *Meaningtwitt is a mashup between Twitter, Google Visualization API and Meaningtool. It is intended for* ***categorization*** *of Twitter Users through the* ***semantic technology*** *of Meaningtool*
[34] *Twitter Sentiment* ***visualizes*** *positive and negative* ***sentiments*** *on Twitter via Google Chart API.*
[35] *Klezio* ***aggregates news*** *from all around the web, provides related stories, wikinews, tags, wikipedia links, flickr photos , tweets, ...*
[36] *MashMeUp creates a* ***data-fused recommendation based on unstructured text****, URLs, or other types of feeds. Also* ***classifies*** *Twitter unstructured information and maps the results across the twitter-verse, applying meaning to twitter*
[37] *SiSense provides common-sense* ***business intelligence software*** *that assists in* ***visualization****,* ***analysis****,* ***dashboarding****, and* ***reporting***
[38] *Newstaggr* ***gathers news*** *from sources like Reuters, CNN, New York Times, and BBC, and it uses Calais to* ***semantically*** *tag them and create a tag cloud for 12 categories*
[39] *Feedz is a feeds aggregator organized by topic where posts are* ***auto-tagged with semantic terms****. Post, share via Twitter or mail, vote, bookmark and create search or* ***semantic alerts***
[40] ***Business Intelligence application*** *from Information Builders. Integrates Google Maps for data visualization*
[41] *VisiStat provides* ***real-time web analytics****. They use the Jigsaw API for LeadCaster, a suite of real-time tools that include the* ***ability to identify companies*** *that visit a website and get their business address, phone numbers, email addresses, etc.*
[42] *Stockaholic uses the EditGrid dynamic images to* ***mashup charts*** *from* ***Yahoo Finance***

Figure 6a Mise en relief des APIs les plus combinées par les mashups pour la veille (centralité de degré >1), les liens avec les mashups ainsi que les mashups ont été effacés du graphe

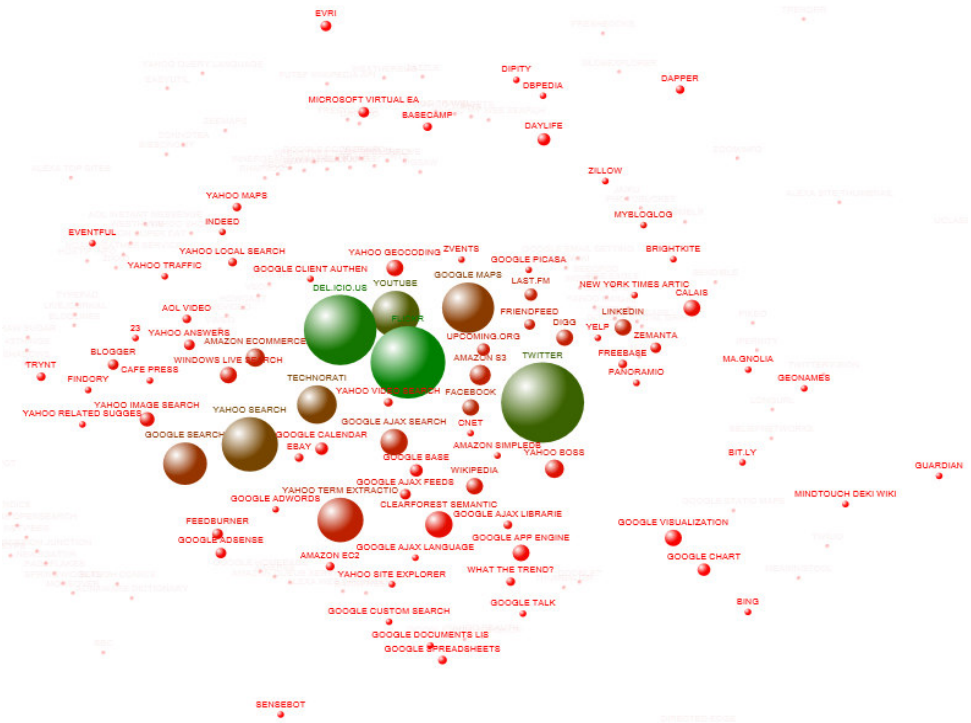


Figure 6b Réseau de relation entre mashups et APIs où Flickr et Youtube ont été enlevés (centralité de degré >4)

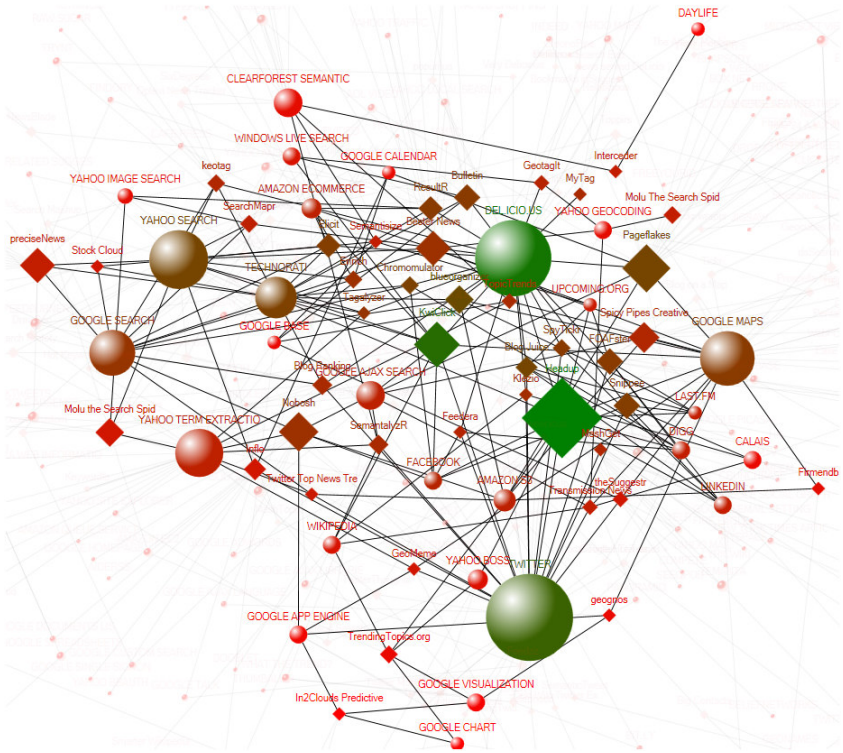


Figure 6 Représentation du réseau APIs (sphères) et mashups (losange) pour la veille. La taille des nœuds est associée à la centralité de degré, la couleur est reliée à la centralité de prestige de l'indice Eigenvector (*les plus prestigieux sont verts jusqu'aux moins prestigieux en rouge), le layout graphique étant celui d'Harel-Koren où les nœuds sont assemblés par similarité*

Nous retrouvons au centre les APIs proposant l'accès à des données brutes, comme évidemment les textes courts et en live de Twitter (avec 35 mashups l'utilisant dans notre corpus), les vidéos de Youtube, les images de Flickr et les bookmarks sociaux de Delicious ou Technorati. Ces données sont aussi souvent combinées avec des APIs de fonctionnalités comme GoogleMaps pour la cartographie et la géolocalisation de données, Google Search et Yahoo Search pour personnaliser des extractions et enfin parmi les plus combinées, notons la présence marquée d'APIs de datamining comme l'API Yahoo Term Extraction (avec 21 mashups l'utilisant dans notre corpus), Yahoo BOSS – Build your One Search Service (relié à des mashups presque tous intéressants pour la veille comme TrendingTopics.org, SenseBot, NewsTrendz, termCloud Search ou LinkSensor) et bien que de plus petite taille et excentrée les APIs Clearforest Semantic Web Services (en lien avec 12 mashups) et le service d'extraction de termes sémantiques Calais de Reuters (en lien avec 7 mashups). Notons qu'aujourd'hui ces deux APIs n'en forment plus qu'une car ClearForest a été rachetée par Reuters en 2009, et intégrée dans l'API OpenCalais, mais elle ressort en tant que telle puisqu'elle est citée dans notre corpus par des mashups plus anciens comme SemmantlizR ou Interceder.

Bien que ces APIs de taille importante et centrales peuvent être considérées comme des plateformes de web services utiles pour la veille, nous trouvons aussi à la périphérie de ce réseau des APIs aux fonctionnalités ciblées pour répondre à un besoin spécifique de veille comme par exemple Evri, Dapper ou Sensebot (services de recherche, d'extraction sémantique de termes et de recombinaison sous diverses formes des données extraites), Dipity (service de visualisation de données sous forme d'une timeline) ou la plateforme ouverte d'accès aux sources du magazine the Guardian[43] (nous avons aussi dans ce même corpus l'équivalent pour le magazine New York Times avec son API New York Times Article Search[44]). La figure 6b présente quant à elle les relations entre Mashups et APIs, les losanges au centre et de grande taille représentant les mashups combinant le plus d'APIs, parmi lesquels nous trouvons des mashups intéressants mixant plus de 10 APIs comme Headup[45] et KwiClick[46], extensions d'extraction de contenu pour Firefox, PageFlake[47], application de construction d'une page d'accueil complètement personnalisée avec du contenu provenant d'autres sources, ou encore le mashup de méta recherche Molu the search spider ou SemantalyzR qui analyse les mots-clés des pages web.

Finalement, l'objectif de la présente étude étant de recenser les combinaisons de mashups et d'APIs incontournables mais aussi singulières pour répondre à des besoins de veille, nous pensons avoir trouvé les réponses à notre questionnement dans cette dernière matrice mashups/APIs et l'analyse de sa représentation graphique[48]. Le classement par la centralité de degré des nœuds Mashups/APIs de notre sélection nous permet, d'une part, en regardant le haut du classement, de repérer les combinaisons les plus évidentes (et donc déjà utilisés par les internautes) et d'autre part, en s'intéressant au bas du classement, de mettre en lumière certaines combinaisons singulières et peut-être innovantes.

Pour résumer, nous avons en haut du classement :
- les APIs de données brutes provenant de sites comme Twitter pour le microbloging, Flickr pour les images (en dans une moindre mesure, Google Picasa et Panoramio), Youtube pour les vidéos, les bookmarks sociaux avec Delicious et Technorati, les réseaux de personnes et sociaux avec Facebook, Lindedin et aussi Friendfeed (à mi-chemin entre un service de bookmark social et un service de réseau social que nous retrouvons dans les mashups précédemment cités KwiClick et FOAFster), les événements avec les sites upcoming.org et Zvents, les sites de recommandations et de reviews comme CNET et DIGG, des sites de fils RSS comme Feedburner et Blogger, de produits d'achats avec les nombreuses APIs d'Amazon (ie Amazon A9 Opensearch, Amazon Ec2, Amazon Ecommerce, Amazon S3 et Amazon Simpledb) ou provenant de Cafe Press et enfin des sites de news comme CrunchBase (l'API du magazine en ligne TechCrunch), celle du New York Times ou de Daylife.
- les APIs de traitements de ces données, centrés autour de termes, de liens, de fils RSS, des méta-données sémantiques et de rendus graphiques. Outre les APIs incontournables de recherche, de cartographique, de visualisation graphique proposées par Google et Yahoo[49], nous trouvons dans cette liste deux autres services de Yahoo Yahoo Term Extraction et Eagle Fire (API de partage de données géolocalisées), Yelp (d'extraction géolocalisée de données financières et de communiqués d'entreprises), Zemanta (API de recherche de données similaires), Calais avec Clearforest Semantic WebServices (le service de dataming sémantique OpenCalais de Thomson Reuters).

---

[43] Cf. http://www.guardian.co.uk/open-platform
[44] Cf. http://developer.nytimes.com/docs/article_search_api
[45] *Headup is a **Semantic Web Firefox addon** that **brings** you **content** related to your interests, friends and contacts.*
[46] *KwiClick is a browser addon that **enhances the way you view and retrieve information** from your favorite services. KwiClick removes the need to open a new tab to get information from Google, Google Maps, Wikipedia, Amazon, YouTube, FriendFeed and more.*
[47] *Custom homepage builder that uses Ajax to let a user layout and customize a page with relevant information, includes pre-built modules for Flickr, YouTube, Amazon and hundreds of other sites.*
[48] Voir Annexe 7.5 pour une représentation complète du réseau Mashups/APIs
[49] i.e. Yahoo Answers, Yahoo Boss, Yahoo Image/Video Search, Yahoo Related Suggestions, Yahoo Search, Yahoo Maps, Yahoo Geocoding ou Google App Engine, Google Calendar, Google Chart, Google Desktop, Google Documents List, Google Homepage, Google Maps, Google Search, Google Social Graph, Google Spreadsheets, Google Talk, Google Visualization, etc…

Les nœuds de plus faible centralité de degré[50] nous mènent sur d'autres APIs, peu mises en relief dans nos analyses précédentes puisque reliées à quelques mashups, mais qui, par le jeu du calcul de popularité et par la représentation graphique se trouvent être singularisées en périphérie du réseau. Ainsi nous avons aussi pu détecter des agencements d'APIs fonctionnant en binôme l'une portant sur l'accès à des données et l'autre sur des fonctionnalités d'analyse comme avec le mashup News Trends Trendrr qui combine les APIs Daylife (Service de news en ligne) et Trendrr (service de représentation graphique de tendances), Newsline mix de Daylife avec Dipity (service de visualisation de données sous forme de timeline), Tickr mix de Dipity, GoogleMap et Flickr qui permet de situer dans le temps et dans l'espace des tags, des utilisateurs et des images de Flickr ou encore The Summarizer mix de Summize Twitter Search, une API de recherche ciblée dans la base de Twitter avec Evri, le service de recherche sémantique de recommandations, pour connecter le monde du microbloging instantané avec d'autres sources d'informations au moyen d'analyse sémantique des métadonnées extraites.

De la même façon, nous pouvons aussi trouver en bas du classement des APIs qui n'ont pas été détectées avec les analyses précédentes (de centralité de degré trop faibles, écrasées dans la masse des données à traiter), et dont les fonctionnalités peuvent tout à fait outiller le veilleur comme Sensebot (service d'extraction sémantique de concept, également installé dans certains services d'Amazon), Beliefnetworks, tout comme Meaningtool ou Direct Edge déclinant des services de recommandation de textes, de site ou de liens basés sur de l'analyse sémantique. Une combinaison particulièrement singulière nous a attiré dans la détection du mashup MDG Actors, dont les créateurs ont obtenu le prix Imagine Cup 2009 de Microsoft, dans le but de trouver un nouveau mode de détection et d'agrégation d'informations sur les acteurs du web apportant leur contribution au bien-être du monde ; en combinant techniquement des news provenant de Google News Twitter et Daylife, avec les APIs Calais et Alchemy d'extraction sémantique de termes, Uclassify système de catégorisation de texte, le tout mise en forme avec l'API de Microsoft PopFly pour assurer la mise en forme de l'analyse des sentiments et des tendances.

Nous avons pu mettre en évidence dans ce chapitre des mashups et APIs connus et moins connus pouvant répondre à chaque étape du processus de veille et qu'il serait intéressant soit de surveiller soit d'intégrer aux prototypes développés dans le cadre du projet de recherche ISICIL. Nous présentons ci-après la répartition des web services détectés au cours de notre démarche exploratoire déployés autour des cinq blocs fonctionnels du processus de veille. Il est intéressant de noter la prépondérance des mashups et APIs sur les trois branches Acquisition, Analytique et Factoring/Visualisation, alors qu'on n'en trouve peu ou pas sur les questions de soutien à la collaboration, au partage, à l'évaluation et plus généralement aux tâches collectives. Il est clair que c'est sur ces points clés que les fournisseurs de plateformes de création de mashups se positionnent dès aujourd'hui comme IBM avec sa solution de création de mashups WebSphere, Microsoft avec SharePoint et PopFly, JackBe avec sa plateforme de composition de web services et widgets Presto, ou encore Convertigo ou Serena.

[50] Voir Annexe 7.5 avec la figure mettant les relations de faible degré en évidence

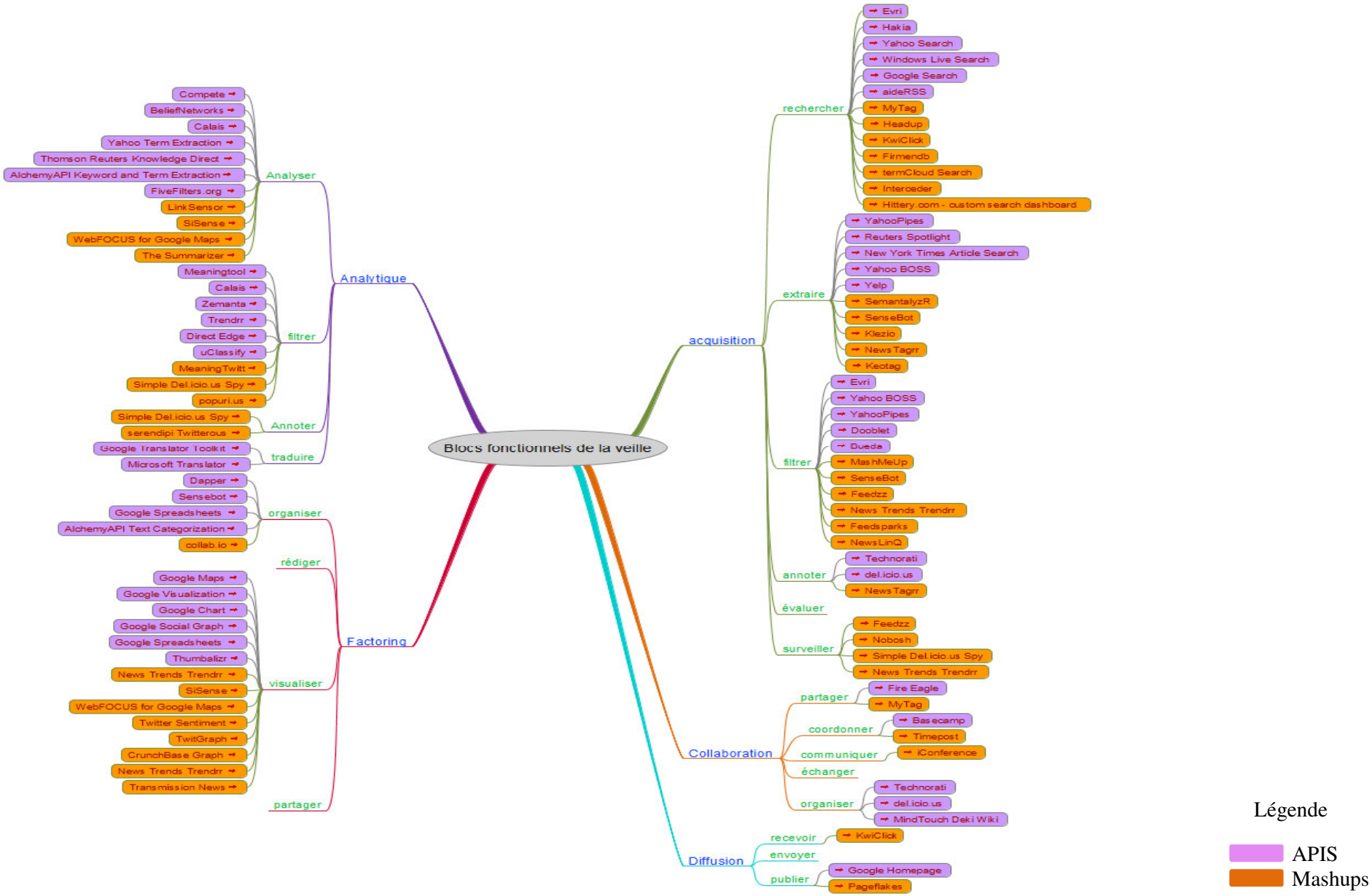


Figure 7 Liste d'APIS et de Mashups pour outiller le processus de veille

5 Conclusion et perspectives

5.1 Les mashups : un marché pour la veille

Nous avons suivi dans cette étude une démarche exploratoire et itérative d'analyse de sources de données complémentaires sur les mashups, pour avoir d'une part une vue amont sur la recherche actuelle et à venir grâce à l'analyse des publications scientifiques spécifiquement sur la question des mashups et une vue aval par l'analyse du marché des mashups sous le prisme du répertoire en ligne ProgrammableWeb. Notre objectif de recherche était, à travers ces analyses, de trouver qu'il existait effectivement des APIs, mashups, widgets et autres web services pouvant être utilisés dans les contextes de veille, de gestion des connaissances et de Business Intelligence.

Nous avons pu démontrer tant dans la littérature que dans la réalité du marché que les mashups, en combinant un éventail large de données augmentées de fonctionnalités de traitement et de visualisation en tout genre, offrent aux travailleurs du savoir des opportunités énormes pour créer des services répondant à leurs besoins de recherche, d'analyse, de mise en forme, de partage et de diffusion d'informations. Les mashups nous semblent représenter La technologie idoine du concept de veille 2.0, remettant le pouvoir à l'analyste-décideur, et rendant possible, dans sa dimension sociale et virale du 2.0, la création d'un espace d'intelligence collective de construction-production de connaissances [44].Vint Cerf[51] lui-même, gourou de l'internet, à propos de la stratégie de Google sur ce marché, soulignait l'énorme potentiel de créativité des internautes, dépassant largement ce qu'une société pourtant leader de l'internet comme Google pouvait espérer produire.

Certes, nous avons aussi vu qu'il existait encore des freins importants à la propagation de ces solutions en entreprise comme la question de l'interopérabilité des services et la garantie de qualité de service bout en bout, la monétisation des mashups lors de leur réutilisation ou encore l'intégrité et la sécurité des données internes de l'entreprise et externes provenant du web. Leur résolution n'est cependant qu'une question de temps, car déjà tout un pan de la Business Intelligence internalise un écosystème de web-services dans les applications d'entreprise déployées (cf. études du Forrester et du Gartner). Le marché des mashups est aussi en train de se structurer ; quand des sociétés parmi lesquels les startups JackBe, Kapow, Convertigo, Dreamface, ou les fournisseurs de technologies comme Intel, HP, Adobe ou encore des sociétés d'autres horizon comme CapGemini, Bank Of America ou Synteractive, se mobilisent dans la création d'un consortium l'Open Mashup Alliance (OMA), dédié à l'utilisation des technologies de Mashup d'entreprises et l'adoption d'un langage ouvert favorisant leur interopérabilité et leur portabilité.

5.2 Perspectives de recherche et approfondissements

Cette étude ouvre également quelques voies qui mériteraient d'être approfondies par exemple en élargissant le spectre des sources de départ à d'autres répertoires de mashups et d'APIs (par exemple Webmashup) et peut-être plus fiables en terme de maintien des liens et d'actualité que Programmableweb, mais aussi dans l'exploration d'autres métriques pour systématiser la mise en corrélation de combinaison d'APIs et notamment pour détecter les combinaisons innovantes (cf. les travaux de Yu sur la métrique des petits mondes – *Small World algorithm*). Il nous a fallu extraire des sous-ensembles de notre corpus initial pour réduire le bruit des nœuds les plus faibles, il serait intéressant de partir du chemin inverse et trouver si, en partant de la liste complète des plus de 2000 APIs de ProgrammableWeb, nous puissions trouver des combinaisons innovantes, car comme l'API Alchemy certaines APIs n'ont pas de mashups qui leur sont reliées.

---

[51] *"What's happening here is the **aggregation of a remarkable collection of people**, all of whom have a very visceral and strong appreciation for what is possible to do with software and information. And they are exploring a variety of ways in which to make these computer-driven tools more useful and also more cross-functional. The focus isn't simply on search. The focus is on **making information discoverable and useful**, so all of these things you see happening at Google are side effects of expanding on the original paradigm, which was making search an effective tool... We know we don't have **a corner on creativity**. There are creative people all around the world, hundreds of millions of them, and they are going to think of things to do with our basic platform that we didn't think of. So the mashup stuff is a wonderful way of allowing people to find new ways of applying the basic infrastructures we're propagating."*
Q&A: Vint Cerf on Google's challenges, aspirations By Juan Carlos Perez November 25, 2005 http://www.computerworld.com/s/article/106535/Q_A_Vint_Cerf_on_Google_s_challenges_aspirations

Nous savons aussi que, pour réaliser cette étude, nous avons mobilisé des méthodes d'analyse quantitative sur deux corpus délimités dans le temps et qu'il faut en mesurer les limites en termes d'interprétation et de projection. Il s'agira pour notre part de poursuivre l'analyse sur le terrain et de confronter les résultats de cette première recherche exploratoire avec les usages, choix, attentes et besoins des communautés de veilleurs et d'évaluer comment de tels services sont adoptés dans le travail au quotidien, au cours de la troisième année du projet ISICIL dédiée aux expérimentations des web services de veille prototypés.

Remerciements

Je remercie tout d'abord l'ANR pour le financement du projet ISICIL ANR-08-CORD-011 qui permet la production de certains éléments mentionnés dans cet article. Merci également pour leurs commentaires constructifs à Eddie Soulier (UTT-TechCICO) et Frédéric Herledan (Orange R&D), relecteurs de la présente étude/livrable.

6 Bibliographie

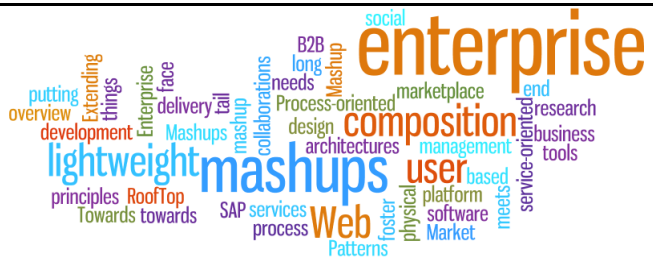
1. IDC. *Worldwide Business Intelligence Tools Vendor Share.* s.l. : IDC #207422E, Volume: 1 Special Reports: Excerpt, 2007.
2. Young, Oliver. *The Mashup Opportunity How To Make Money In The Evolving Mashup Ecosystem.* s.l. : Forrester Research, 2008.
3. Schlegel, K. *Predicts 2009: Business Intelligence and Performance Management Will Deliver Greater Business Value* . s.l. : Gartner Group, 2008.
4. Rojas, Pete. Bootleg Culture. *Salon.com.* [Online] 2002. http://dir.salon.com/story/tech/feature/2002/08/01/bootlegs/index.html.
5. Berinstein, P. Location, location, location: online maps for the masses. *Searcher - Information Today.* jan 2006, pp. vol.14, no.1, pp. 16-25.
6. Weinberger, D. *If you love your information, set it free.* 2007, Harvard Business Review, pp. 85, 6, pp 20–2.
7. Merrill, Duane. *Mashups: The new breed of Web app. An introduction to mashups.* s.l. : IBM developerWorks, 2006.
8. Hinchcliffe, Dion. *An Executive Guide to Mashups in the Enterprise.* s.l. : JackBe - Hinchcliffe & Company, 2008.
9. Fagan, J.C. Using Yahoo! Pipes. *Computers in Libraries.* Information Today Inc, 2007, Vols. Nov.-Dec. 2007, vol.27, no.10, pp. 11-12, 14, 16-17.
10. Vriens, D. The role of information and communication technology in competitive intelligence. *Information and Communication Technology.* Hershey : Idea Group Publishing, 2004, p. 305 pp.
11. Benczúr, D. *Environmental scanning: how developed is information acquisition in Western European companies?* 2005, Information Research, pp. Oct 2005, Vol. 11 Issue 1, p1-1, 1p.
12. Leitzelman, M. La veille 2.0 : Outiller les interactions sociales au sein du processus de veille. *Les Cahiers du numérique : Du web 2.0 au concept 2.0.* janv 2010, pp. Volume 6, p200.
13. Kobielus, J. *Mighty Mashups: Do-It-Yourself Business Intelligence For The New Economy.* s.l. : Forrester Institute, 2009.
14. Alba, A.&al. *Applications of voting theory to information mashups*Second IEEE International Conference on Semantic Computing (ICSC), 2008. pp. 10-17.
15. Li, Chung-Sheng *Enterprise information mashup and real time business assurance for global collaboration.*. Atlanta : s.n., 2006. International Conference on Collaborative Computing: Networking, Applications and Worksharing. pp. pp. 514 - 536.
16. Maximilien, E.M. *Web services on Rails: using Ruby and Rails for Web services development and mashups.* Chicago : s.n., 2006. 2006 IEEE International Conference on Web Services. p. 1 pp.
17. Mulholland, A. *The end of business as usual: service-oriented business transformation.* Chicago : s.n., 2006. Service-Oriented Computing - ICSOC 2006. 4th International Conference. p. Lecture Notes in Computer Science Vol. 4294.
18. MacVittie, L. *Danger 2.0 [Web 2.0 technologies].* 2006, Network Computing, pp. vol.17, no.22, pp. 40-6, 48, 51-2, 54.
19. Ayhan, S., Comitz, P. and Stemkovski, V. *Aviation Mashups.* Orlando : s.n., 2009. IEEE/AIAA 28th Digital Avionics Systems Conference. DASC 2009. p. 5 pp.

20. Gurram, R., Mo, B. and Gueldemeister, R. *A Web based mashup platform for enterprise 2.0.* Auckland, New Zealand : s.n., 2008. Web Information Systems Engineering - WISE 2008 International Workshops.
21. Muñoz, J, Hernandez, JR and Ninova, M. *A Network Analysis of the Term "Social Networks" in the Context of ICT Academic Literature.* Zurich : s.n., 2007. ASNA - Applications of Social Network Analysis.
22. Gultom, R.A.G., Sari, R.F. and Budiardjo, B. *Implementing web data extraction and making Mashup with Xtractorz*Patiala, India : IEEE, Piscataway, NJ, USA, 2010. IEEE 2nd International Advance Computing Conference (IACC 2010). pp. pp. 385-93.
23. Sari, R.F. and Kurniawan, A. *Implementation of Indonesian electronic Citation System Based on Web Extraction Techniques*Phuket, Thailand : s.n., 2010. 3rd International Conference on Knowledge Discovery and Data Mining (WKDD 2010). pp. pp. 494-7, xiv+610 pp.
24. Ngu, A.H.H., &al. *Semantic-Based Mashup of Composite Applications.* 2010, IEEE Transactions on Services Computing, pp. Jan.-March 2010, vol.3, no.1, pp. 2-15.
25. Toubiana, V., et al. *R2M: a reputation model for mashups*Las Vegas, NV, USA : s.n., 2010. 7th IEEE Consumer Communications and Networking Conference (CCNC). p. 6 pp.
26. Klym, N., Montpetit, M.J. and Blain, E.*Building social services.* Las Vegas, NV, USA : s.n., 2010. 7th IEEE Consumer Communications and Networking Conference (CCNC). p. 5 pp.
27. Chen, R., et al.*Applying SOA and Web 2.0 to telecom: legacy and IMS next-generation architectures.* Xi'an, China : s.n., 2008. IEEE International Conference on e-Business Engineering. pp. pp. 374-9.
28. Zou, J. and Pavlovski, C.J. *Towards accountable enterprise mashup services.* Hong Kong, China : s.n., 2007. IEEE International Conference on e-Business Engineering. pp. pp. 205-12.
29. Trapero Burgos, R., et al. *Next generation mashups: how to create my own services in a convergent world.* 2009, IEEE Latin America Transactions, pp. vol.7, no.3, pp. 390-4.
30. Nakano, Y., Yamato, Y. and Sunaga, H. *Web-service-based avatar service modelling in the next generation networ.* Bandos Island, Maldives : s.n., 2008. 7th Asia-Pacific Symposium on Information and Telecommunication Technologies. pp. pp. 52-7.
31. Wang, Shuangquan, Chen, Canfeng and Ma, Jian. *A framework for wireless sensor network based mobile mashup applications.* Los Angeles, CA, USA : s.n., 2009. WRI World Congress on Computer Science and Information Engineering, CSIE. pp. vol.5, pp. 683-6.
32. Brule, M. and Hodges, C. *Improving E&P performance with right-time business intelligence.* 2007, World Oil, pp. Nov. 2007, vol.228, no.11, pp. 83-8.
33. Pirolli, P. and Card, S. *Sensemaking Processes of Intelligence Analysts and Possible Leverage Points as Identified Through Cognitive Task Analysis*McLean, USA : https://analysis.mitre.org/proceedings/Final_Papers_Files/206_Camera_Ready_Paper.pdf, 2005. International Conference on Intelligence Analysis.
34. Kuhlthau, C. C. *Seeking meaning: a process approach to library and information services.* 2nd. ed. Westport : Libraries Unlimited, 2004.
35. Chatti, M.A., Jarke, M. and Specht, M. *PLEF: a conceptual framework for mashup personal learning environments*2009, IEEE Learning Technology, pp. July 2009, vol.11, no.3, pp. 20-3.
36. Modritscher, F. and Wild, F.*Personalized e-learning through environment design and collaborative activities* Graz, Austria : s.n., 2008. HCI and Usability for Education and Work. 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008. pp. pp. 377-90.
37. Dong, Bo, et al. *Jampots: a mashup system towards an E-learning ecosystem*Seoul, South Korea : s.n., 2009. Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC. pp. pp. 200-5.
38. Borau, K., Ullrich, C. and Kroop, S. *Mobile learning with Open Learning Environments at Shanghai Jiao Tong University, China.* 2009, IEEE Learning Technology, pp. Jan.-April 2009, vol.11, no.1-2, pp. 7-9.
39. Meza, J. and Zhu, Qin. *Mix, match, rediscovery: a mashup experiment of knowledge organization in an enterprise environment.* 2008, International Journal of Knowledge Management, pp. Jan.-March 2008, vol.4, no.1, pp. 65-76.

40. Von Hippel, E. *Democratizing Innovation.* s.l. : MIT Press, 2005, Vols. pp 93-106.
41. Yu, S and Woodard, C.J. *Innovation in the Programmable Web: Characterizing the Mashup Ecosystem.* s.l. : Lecture Notes in Computer Science 5472 Springer 2009, 2008. Second International Workshop on Web APIs and Services Mashups (Mashups 2008).
42. Smith, M. A., et al. *Analyzing (social media) networks with NodeXL.* New York : s.n., 2009. Proceedings of the Fourth international Conference on Communities and Technologies C&T '09. ACM. pp. 255-264 pp.
43. Perer, A. and Shneiderman, B.*Systematic yet flexible discovery: guiding domain experts through exploratory data analysis.* Gran Canaria, Spain : s.n., 2008. Proceedings of the 13th international Conference on intelligent User interfaces. pp. 109-118 pp.
44. Bulinge, F. *Le cycle du renseignement : analyse critique d'un modèle empirique.* 2006, Market Management, p. pp 36 à 52.
45. Jagadeesan, L.J. and Mendiratta, V.B. *Voice communication mashups: formal specification and composition of service level agreements*Cardiff, Wales, UK : s.n., 2009. Third International Conference on Next Generation Mobile Applications, Services and Technologies. NGMAST 2009. pp. pp. 131-6.
46. Wenger, E, et al. Technology for communities. [book auth.] E Wenger, N White and JD Smith. *Digital Habitats: stewarding technology for communities.* Portland : CPsquare, 2009, p. 218 pp.
47. Koschmider, A, Torres, V and Pelechano, V. *Elucidating the Mashup Hype: Definitions, Challenges, Methodical Guide and Tools for Mashups.* Madrid, Spain : s.n., 2009. 2nd Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web (MEM 2009).
48. Hoyer, V., and Fischer, M. *Market Overview of Enterprise Mashup Tools. In International Conference on Service oriented Computing.* 2008, Volume 5364 of Lecture Notes in Computer Science, pp. 708-721 pp.
49. Junseok, H, Altmann, J and K, Kibae. *The Structural Evolution of the Web2.0 Service Network.* Seoul : SEOUL NATIONAL UNIVERSITY Discussion Paper No. 2009:14, 2009.
50. Zhang, Libin, et al. *Web 2.0 and semantic Web based intelligent service system for tourism.* Dalian, China : s.n., 2008. 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM).
51. Leitzelman, M. *Etat de l'art et tendances sur le marché de la veille et de l'Intelligence compétitive : Positionnement théorique, analyse du marché, approche fonctionnelle du processus de veille.* s.l. : ANR-08-CORD-011-05, 2009.
52. West, J. Does appropriability enable or retard open innovation. [book auth.] H Chesbrough, W Vanhaverbeke and J West. *Open innovation: Researching a new paradigm.* s.l. : Oxford University Press, 2006.

7 Annexes

7.1 Tagclouds des sociétés publiantes (extraction des mots-clés du champ KW)

Société	Publi.	Mots-clés les plus représentatifs	Société	Publi.	Mots-clés les plus représentatifs
IBM	23		Hewlett Packard	5	
SAP	9		Sun Microsystems	5	
Nokia	8		Telefonica Research & Development	4	
Alcatel Lucent	6		Microsoft	3	

7.2 Liste des conférences 2010-2009

Année	Pays	Titre de la conférence	Pub	Année	Pays	Titre de la conférence	Pub
2010	Thailand	2010 3rd International Conference on Knowledge Discovery and Data Mining (WKDD 2010)	1	2009	Greece	2009 International Conference on Advances in Social Network Analysis and Mining (ASONAM)	1
2010	NV, USA	2010 7th IEEE Consumer Communications and Networking Conference (CCNC)	2	2009	Japan	2009 International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2009)	1
2010	India	2010 IEEE 2nd International Advance Computing Conference (IACC 2010)	1	2009	NV, USA	2009 International Conference on Consumer Electronics (ICCE)	2
2010	Netherlands Antilles	2010 Second International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2010)	1	2009	UK	2009 International Conference on CyberWorlds (CW)	1
2009	South Korea	2009 11th International Conference on Advanced Communication Technology	1	2009	China	2009 International Conference on Information Management, Innovation Management and Industrial Engineering	1
2009	France	2009 13th International Conference on Intelligence in Next Generation Networks (ICIN): 'Beyond the Bit Pipes'	3	2009	China	2009 International Conference on Management and Service Science (MASS)	2
2009	VA, USA	2009 17th International Conference on Geoinformatics	1	2009	China	2009 International Conference on Management of e-Commerce and e-Government (ICMECG)	1
2009	Netherlands	2009 18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE)	1	2009	China	2009 International Conference on New Trends in Information and Service Science (NISS 2009)	1
2009	Cyprus	2009 24th International Symposium on Computer and Information Sciences (ISCIS)	1	2009	United Arab Emirates	2009 International Conference on the Current Trends in Information Technology (CTIT 2009)	1
2009	China	2009 2nd IEEE International Conference on Computer Science and Information Technology	1	2009	MD, USA	2009 International Symposium on Collaborative Technologies and Systems (CTS 2009)	1
2009	Greece	2009 35th Euromicro Conference on Software Engineering and Advanced Applications	1	2009	South Korea	2009 International Symposium on Ubiquitous Virtual Reality (ISUVR 2009)	1
2009	France	2009 3rd IEEE International Conference on Digital Ecosystems and Technologies (DEST)	3	2009	Taiwan	2009 Joint Conferences on Pervasive Computing (JCPC 2009)	1
2009	India	2009 3rd International Conference on Internet Multimedia Services Architecture and Application (IMSAA)	1	2009	UK	2009 Joint Working IEEE/IFIP Conference on Software Architecture (WICSA) & 3rd European Conference on Software Architecture (ECSA)	1
2009	Spain	2009 4th International Workshop on Semantic Media Adaptation and Personalization (SMAP 2009)	1	2009	Netherlands	2009 Seventh IEEE European Conference on Web Services. ECOWS 2009	1
2009	China	2009 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)	1	2009	Japan	2009 Seventh International Conference on Creating, Connecting and Collaborating through Computing (C5 2009)	1
2009	China	2009 Fifth International Conference on Semantics, Knowledge and Grid (SKG 2009)	1	2009	PA, USA	2009 Sixth International Conference on Networked Sensing Systems (INSS 2009)	1
2009	South Korea	2009 Fifth International Joint Conference on INC, IMS and IDC	1	2009	Malta	2009 Third International Conference on Advances in Semantic Processing (SEMAPRO 2009)	1
2009	Shandong, China	2009 Fourth ChinaGrid Annual Conference. ChinaGrid 2009	1	2009	Wales, UK	2009 Third International Conference on Next Generation Mobile Applications, Services and Technologies. NGMAST	1
2009	South Korea	2009 Fourth International Conference on Computer Sciences and Convergence Information Technology	1	2009	China	2009 Third International Symposium on Intelligent Information Technology Application	1
2009	Japan	2009 ICROS-SICE International Joint Conference. ICCAS-SICE 2009	1	2009	TX, USA	2009 Winter Simulation Conference (WSC 2009)	1
2009	BC, Canada	2009 ICSE Workshop on Principles of Engineering Service Oriented Systems (PESOS 2009)	1	2009	Switzerland	2009 Workshops at the Grid and Pervasive Computing Conference. GPC 2009	1
2009	UK	2009 IEEE 23rd International Conference on Advanced Information Networking and Applications Workshops (WAINA)	1	2009	NB, Canada	2009 World Congress on Privacy, Security, Trust and the Management of e-Business. CONGRESS 2009	1
2009	China	2009 IEEE 25th International Conference on Data Engineering. ICDE 2009	2	2009	CA, USA	2009 WRI World Congress on Computer Science and Information Engineering, CSIE	1
2009	Singapore	2009 IEEE Asia-Pacific Services Computing Conference (APSCC 2009)	2	2009	Austria	E-Commerce and Web Technologies. 10th International Conference, EC-Web 2009	1
2009	Austria	2009 IEEE Conference on Commerce and Enterprise Computing	1	2009	HI, USA	GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference	1
2009	CA, USA	2009 IEEE Congress on Services. SERVICES 2009	3	2009	CA, USA	Human Interface and the Management of Information. Designing Information Environments. Symposium on Human Interface 2009	1
2009	China	2009 IEEE International Conference on e-Business Engineering. ICEBE 2009	2	2009	Sweden	Human-Computer Interaction - INTERACT 2009. 12th IFIP TC 13 International Conference	1
2009	NV, USA	2009 IEEE International Conference on Information Reuse & Integration (IRI 2009)	1	2009	MS, USA	OCEANS 2009	1
2009	India	2009 IEEE International Conference on Services Computing (SCC)	1	2009	Netherlands	Scalability of Networks and Services. Third International Conference on Autonomous Infrastructure, Management and Security, AIMS 2009	1
2009	CA, USA	2009 IEEE International Conference on Web Services (ICWS)	8	2009	LO, USA	Scientific and Statistical Database Management. 21st International Conference, SSDBM 2009	1
2009	China	2009 IEEE International Symposium on IT in Medicine & Education (ITME2009)	1	2009	Austria	Service-Oriented Computing - ICSOC 2007 Workshops	4
2009	MA, USA	2009 IEEE International Symposium on Performance Analysis of Systems and Software	1	2009	NSW, Australia	Service-Oriented Computing. ICSOC 2008 International Workshops	6
2009	AZ, USA	2009 IEEE International Symposium on Sustainable Systems and Technology (ISSST)	1	2009	Russia	Smart Spaces and Next Generation Wired/Wireless Networking. 9th International Conference, NEW2AN 2009 and Second Conference on Smart Spaces, ruSMART	1
2009	Tunisia	2009 IEEE Symposium on Computers and Communications (ISCC)	1	2009	Serbia	TELSIKS 2009 - 2009 9th International Conference on Telecommunications in Modern Satellite, Cable, and Broadcasting Services	1
2009	OR, USA	2009 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)	2	2009	CA, USA	Universal Access in Human-Computer Interaction. Applications and Services. 5th International Conference, UAHCI 2009. Held as Part of HCI International 2009	1
2009	FL, USA	2009 IEEE/AIAA 28th Digital Avionics Systems Conference. DASC 2009	1	2009	Spain	Web Engineering. 9th International Conference, ICWE 2009	4
2009	VA, USA	2009 Integrated Communications, Navigation and Surveillance Conference (ICNS 2009)	1				

7.3 Codes de classification d'INPEC dédiés à la veille

Figure 8a Classement des Codes de classification des articles du corpus (freq.>10) : présence significative des domaines de recherche proches de la veille

Formes	Fréquence
C6150N Distributed systems software	207
C7210N Information networks	133
C6110B Software engineering techniques	59
C7100 Business and administration	46
C6150E General utility programs	27
C6180 User interfaces	25
C6130S Data security	24
C6170K Knowledge engineering techniques	23
B6210L Computer communications	21
C7250N Search engines	20
B6250F Mobile radio systems	17
C7810C Computer-aided instruction	17
C6110J Object-oriented programming	15
D2010 Business and professional IT applications	15
C5620W Other computer networks	13
C6130D Document processing techniques	13
D2080 Information services and database systems in IT	13
C6130G Groupware	12
C6130M Multimedia	12
C6150J Operating systems	12
C7250R Information retrieval techniques	11
D5000 Office automation - computing	11

Figure 8b Présence constante des domaines de recherche proches de la veille (ex le domaine « Knowledge engineering techniques » issu des Codes de Classification d'INPEC concerne 10 articles du corpus en 2009, 10 en 2008 et 3 en 2007)

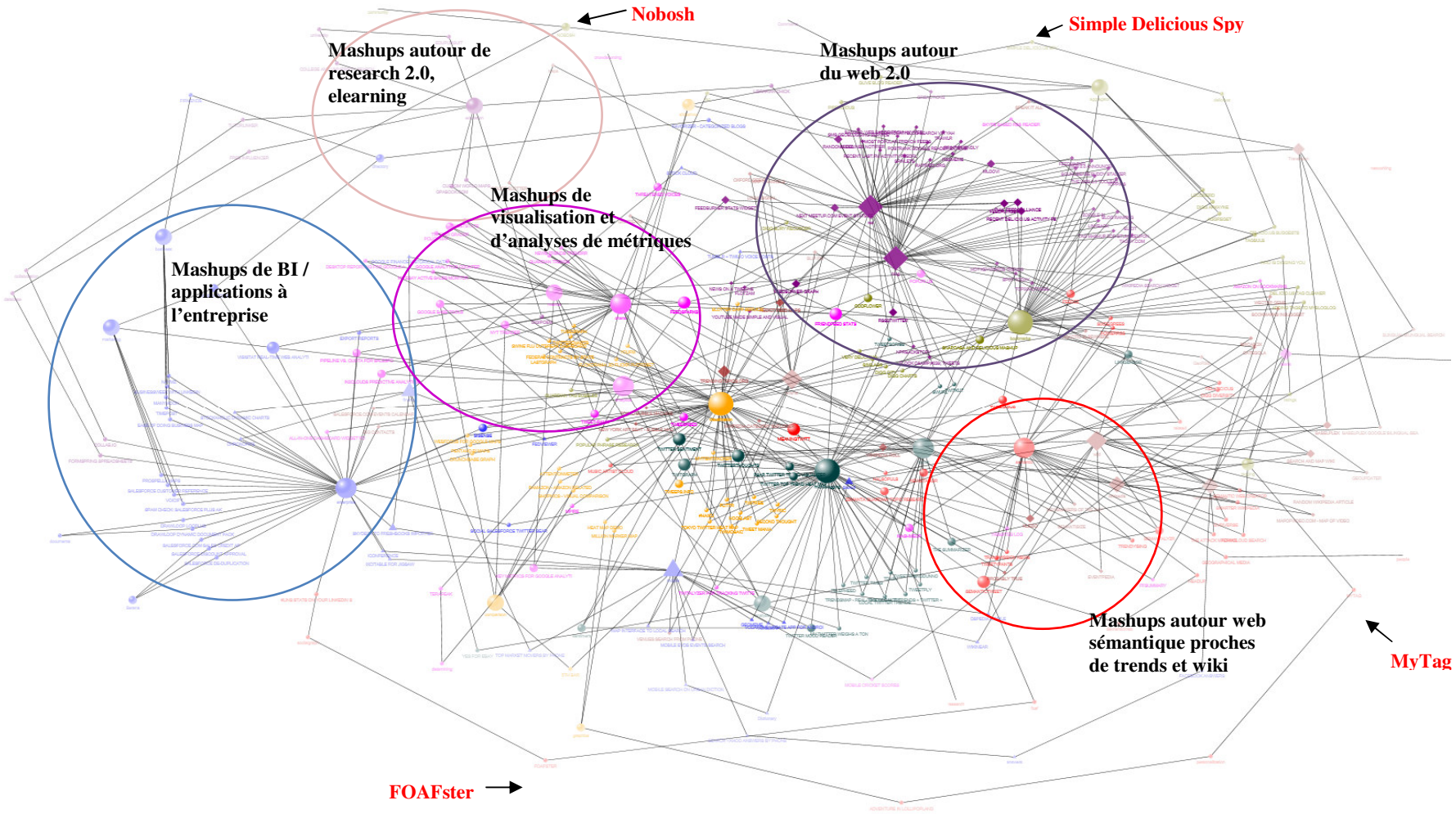
Formes (CCVeille)	Formes (SI)	Fréquence
C7250R Information retrieval techniques	2010	2
C6130 Data handling techniques	2010	1
C7240 Information analysis and indexing	2010	1
C6170K Knowledge engineering techniques	2009	10
C6130D Document processing techniques	2009	6
C6130G Groupware	2009	5
C7250N Search engines	2009	4
D2010 Business and professional IT applications	2009	4
C7220 Generation, dissemination, and use of information	2009	4
D2080 Information services and database systems in IT	2009	3
C6130 Data handling techniques	2009	3
C6120 File organisation	2009	3
C7250R Information retrieval techniques	2009	2
C7820 Humanities computing	2009	2
C7170 Marketing computing	2009	2
D5000 Office automation - computing	2009	2
C7240 Information analysis and indexing	2009	1
C6160N Natural language processing	2009	1
A9365 Data and information; acquisition, processing, storage and dissemination in geophysics	2009	1
C6170K Knowledge engineering techniques	2008	10
C7250N Search engines	2008	8
C7250R Information retrieval techniques	2008	7
C6130D Document processing techniques	2008	6
D2010 Business and professional IT applications	2008	4
C6130G Groupware	2008	4
C6130 Data handling techniques	2008	3
C7820 Humanities computing	2008	3
D5000 Office automation - computing	2008	3
D2080 Information services and database systems in IT	2008	2
C7240 Information analysis and indexing	2008	1
C7170 Marketing computing	2008	1
D1040 Human aspects of IT	2008	1
C7250 Information storage and retrieval	2008	1
D1000 General & Management aspects of Information Technology	2008	1
C7250N Search engines	2007	7
D2010 Business and professional IT applications	2007	6
D2080 Information services and database systems in IT	2007	6
D5000 Office automation - computing	2007	5
C6170K Knowledge engineering techniques	2007	3
C6130G Groupware	2007	3
C7220 Generation, dissemination, and use of information	2007	2
C6130D Document processing techniques	2007	1
C6130 Data handling techniques	2007	1
C7820 Humanities computing	2007	1
E0420 Information management	2007	1
C7250 Information storage and retrieval	2007	1
D2080 Information services and database systems in IT	2006	2
C7250N Search engines	2006	1
D2010 Business and professional IT applications	2006	1
C7820 Humanities computing	2006	1
D5000 Office automation - computing	2006	1

Figure 8 Application claire des mashups dans les champs de recherche proches de la veille

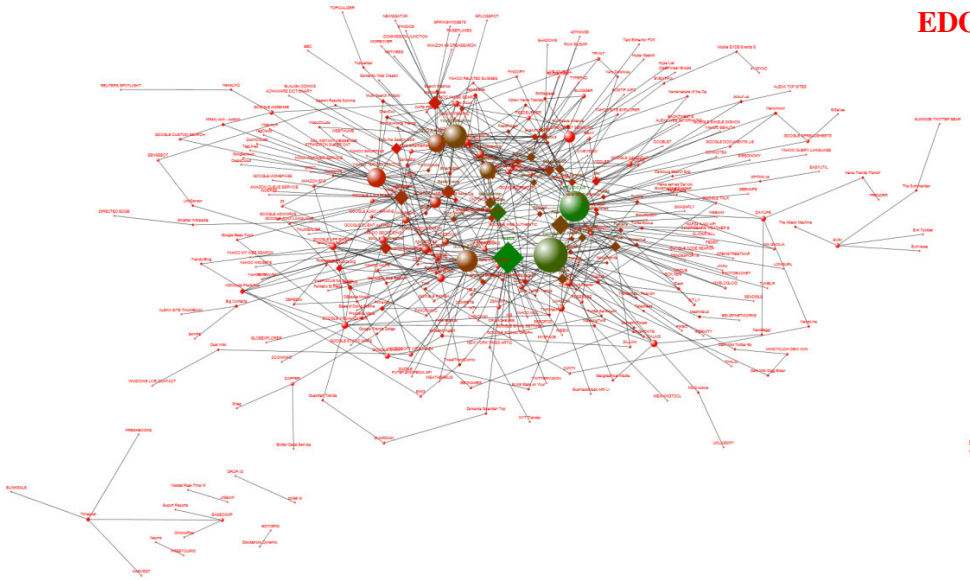
NB : la figure 8b présente détaillée par année et par fréquence une sélection de codes de classification[52] proches des activités de veille, comme par exemple les champs de recherche *Information retrieval techniques*, *Knowledge engineering techniques*, *Search engines*, ou encore *Information analysis and indexing* ; nous pouvons remarquer une fréquence cumulée relativement importante de ces champs de recherche soulignant aussi l'intérêt des mashups pour des applications de veille (cf. Figure 8a vis-à-vis des autres codes de classification).

[52] Il s'agit du champ CC – Classification Code fixé par les indexeurs d'INSPEC

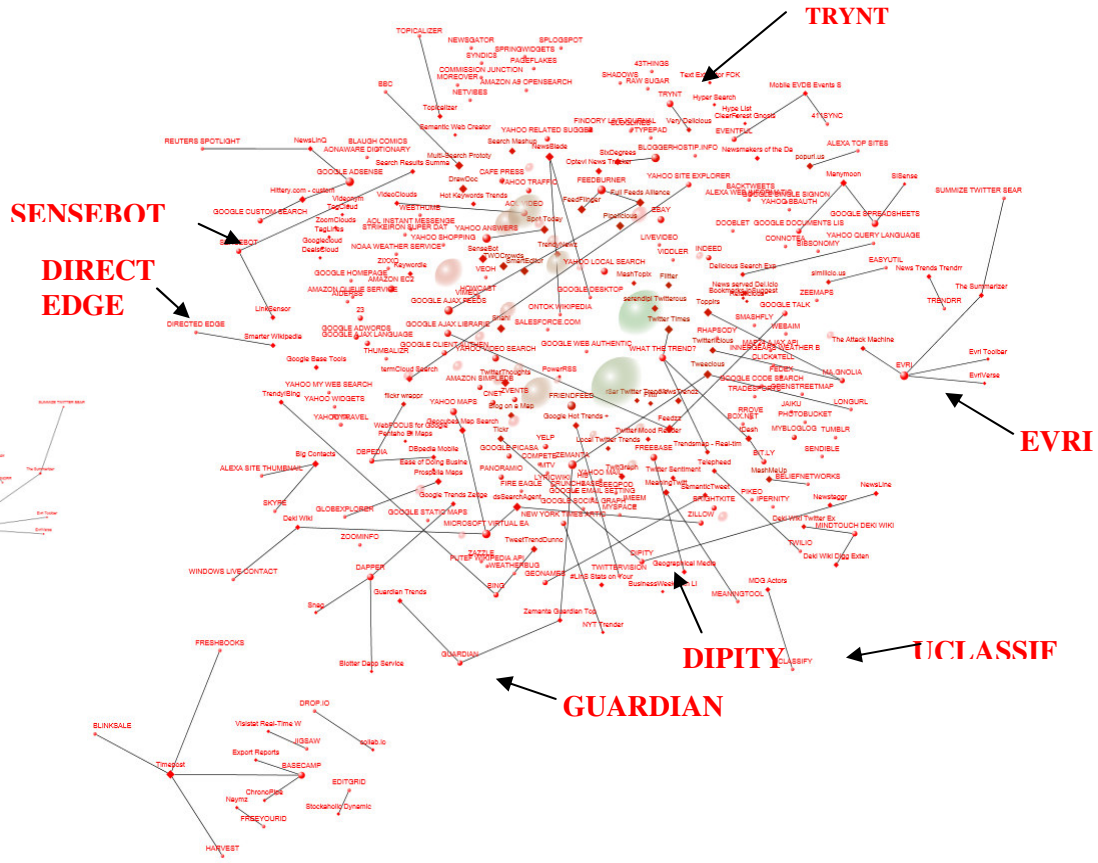
7.4 Visualisation des clusters par famille de mashups autour des tags représentatifs de la veille



7.5 Réseau complet des combinaisons de mashups et APIs pour la veille



Visualisation complète de tous les nœuds



Visualisation des nœuds de faible centralité de degré < 4 (les moins populaires)