Infrastructure Web socio-sémantique pour la Veille Collaborative

Jean-Pierre Cahier, Mylène Leitzelman, Patrick Brébion

▶ **To cite this version:**

Jean-Pierre Cahier, Mylène Leitzelman, Patrick Brébion. Infrastructure Web socio-sémantique pour la Veille Collaborative. Catherine Faron-Zucker. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.267-269. hal-01016080

HAL Id: hal-01016080

https://inria.hal.science/hal-01016080

Submitted on 27 Jun 2014

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

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

Infrastructure Web socio-sémantique pour la Veille Collaborative

Jean-Pierre Cahier[1], Mylène Leitzelman[2] et Patrick Brébion[3]

[1] Université de Technologie de Troyes (UTT) - Equipe ICD/Tech-CICO,
cahier@utt.fr
[2] Société Mnémotix, 06210 Mandelieu
mylene.leitzelman@mnemotix.com
[3] Journaliste spécialisé indépendant,
Pbrebion@yahoo.fr

Résumé : L'activité de veille est de plus en plus organisée et gérée en projet collectif, où les équipes de veilleurs doivent multiplier les interactions avec les réseaux d'utilisateurs, en mettant en jeu une variété d'objets (pages Web, articles, projets, équipes, questionnements,...) qu'il s'agit pour les analystes de d'appréhender dans une variété de points de vue et d'activités (analyse qualitative ancrée dans les documents, le catalogage, la visualisation des réseaux sociaux associés aux thèmes du domaine, l'annotation partagée, la co-écriture de rapports, etc.). Nous proposons dans le domaine de la Veille d'expérimenter et évaluer sur plusieurs terrains l'apport d'une approche innovante d'ingénierie de connaissances collaborative utilisant une infrastructure "Web socio sémantique", réunissant notamment les Webmarks (Delaforge *et al.*, 2012) et une série d'outils gérant des points de vue multiples (Cahier *et al.*, 2013).

Mots-clés : web socio-sémantique, travail collaboratif, Webmarks, points de vue, veille.

En accord avec (Weick, 1995) (Baumard, 1991)(Bulinge, 2006), nous pensons (Leitzelman, 2010) que le cycle classique de la veille formalisé par une approche rationnalisée et itérative de l'information a ses lacunes. L'analyse du processus de veille est restée longtemps centrée sur le rapport de l'individu à l'information, or le web 2.0 et l'adoption massive des technologies sociales, virales et mobiles dans la sphère privée puis dans la sphère professionnelle, ont modifié ces rapports de l'individu au groupe dans le monde du travail.

La veille demande une approche collaborative, que nous proposons de baser sur un ensemble d'outils, partageant une infrastructure commune aux standards du Web. Ces outils supposent des modèles de connaissances et de ressources adaptés au travail de groupe, ce qui les place dans le champ du Web socio-sémantique pour assurer la jonction avec les techniques de Web social (modèles d'utilisateurs...) et de Web sémantique (modèles de concepts.) Notre hypothèse, que nous avons commencé à mettre à l'épreuve sur plusieurs terrains, est que cet ensemble d'outils d'ingénierie de connaissances collaborative, est de nature à faciliter grandement l'activité collective de veille, en permettant de travailler plus rapidement avec une qualité de veille accrue sur des champs informationnels plus complexes.

1 La veille comme activité collaborative

Les activités de veille 2.0 que nous considérons ne consistent pas à étudier un corpus de documents primaires figé au départ, mais au contraire à le constituer de façon incrémentale et permanente en cherchant aussi à enquêter au contact de membres des communautés

concernées et en confrontation avec les débats qui les traversent. Ce qui implique de laisser « la main » aux veilleurs sur la définition des sources et des corpus en temps continu.

Répondre à cette problématique en prenant en compte des volumes de documents à surveiller se traduit par l'utilisation d'outils support facilitant l'identification des données, notamment, le signal faible, sans automatiser les étapes d'affectation de thèmes et de catégories. Cette intervention humaine autorise l'enrichissement dynamique des éléments pour ajuster en continu le rapport entre le bruit et le silence. Elle facilite également l'identification du signal faible. Les objets d'intérêt de veilleurs peuvent être détectés ou émerger dans des lectures de documents mais aussi dans des interactions entre veilleurs (e.g. des tags, des recommandations de lectures, des co-écritures de notices intermédiaires, des évaluations de réputation, des chats…) ou entre veilleurs et acteurs du domaine..

Pour dépasser les limites de l'approche classique, un système de veille doit permettre aux acteurs de réajuster facilement les périmètres et les corpus et de mettre en concurrence et en correspondance plusieurs vues sur les tendances latentes. Les étapes suivantes d'enrichissement et d'analyse se doivent également d'être basées sur la même logique pour profiter de cette dynamique et de cette approche multipoints de vue. Des corpus ainsi que des jeux de catégories d'analyse évolutifs et pluriels, seront ainsi de nature à refléter l'existence de plusieurs discours voire de controverses.

2 Infrastructure " Web socio-sémantique " proposée

L'architecture fonctionnelle de l'infrastructure d'outils que nous proposons d'expérimenter, sans être attachée à un processus ni à une méthode de veille uniques, considère la veille comme l'activité collaborative riche décrite précédemment. L'approche technologique suivie consiste à rendre interopérables et utiliser grâce à une plate-forme en Services Web REST une série d'outils de logiciel libre issus des recherches de nos équipes (dont certains ont été développé ou intégrés par les auteurs) tous ayant fait l'objet par ailleurs de publications.

Un premier enjeu est la cons**t**itution d'un paysage thématique partagé d'un domaine vaste, incluant la délimitation et la structuration graduelles du domaine cible. Un premier aspect de la structuration concerne l'identification progressive des items et des corpus pertinents. Ici cette activité utilise les outils Argos et Agoræ (Zaher et al., 2006).

TABLE 1 – *Principales fonctions utilisées dans les expériences projetées et outils correspondants.*

Principales fonctionnalités mentionnées	**Références des outils et/ou démos**
Bookmarking social et sémantique	Plateforme Webmarks (http://mnemotix.com)
Chaîne d'annotation sémantique	Plateforme Webmarks (http://mnemotix.com)
Analyse des réseaux d'acteurs	Plateforme Webmarks / module SemSNA (http://mnemotix.com)
Middleware sémantique (service JSON)	Mnémokit (http://mnemotix.com)
Cataloguage multi-points de vue	http://hypertopic.org/Knowledge_management.html
Analyse qualitative ancrée dans les documents	http://hypertopic.org/Texts_analysis.html
Analyse des co-occurences items/thèmes	https://github.com/Hypertopic/Porphyry/wiki
Middleware socio-sémantique Hypertopic (REST)	https://github.com/Hypertopic/Protocol/wiki

Le modèle Hypertopic V2 (Zhou, 2006) sous-jacent permet de définir plusieurs types d'items et de les taguer selon plusieurs points de vue, qui peuvent être des opinions, des dimensions d'analyse partagées, ou des catégories importées ou construites par l'analyste avec d'autres outils de la plate-forme proposée. Le réseau de thème et d'items aussi être exploité à l'aide d'un autre outil (Porphyry) procurant alors des fonctions supplémentaires de visualisation fine des co-occurrences entre thèmes, rendant alors visible le réseau de relations entre des thèmes utilisés pour décrire un même item, y compris si ces thèmes appartiennent à des points de vue différents. Cassandre, qui est un outil d'analyse qualitative permettant le travail de thématisation et de catégorisation "manuelle" procure par ailleurs aux veilleurs une

aide semi-automatisée en faisant apparaitre, si besoin, des comptages lexicométriques des mots ou groupes spécifiques ou répétés (ou au contraire rares) (Lejeune & Bénel, 2012) pour détecter des régularités ou des singularités dans une lecture de survol. Grâce à l'usage conjugué de l'outil LaSuli (Bénel et al., 2010), organisé tout comme Cassandre par le modèle Hypertopic, les fragments peuvent être tagués selon des points de vue et reliés ainsi à des catégories d'analyse sous contrôle total du lecteur, l'analyste pouvant faire apparaitre une catégorie avant de lui donner un nom, ce qui est classique dans une démarche d'enquête.

Une fonctionnalité de wiki sémantique, incluse également dans la plate-forme proposée, a été développée et déployée dans le but double d'assurer l'édition collaborative de documents d'une part, et en corollaire, de capturer des traces d'activités permettant d'inférer des relations sociales au sein des membres de la communauté de veille testée (Buffa et al., 2012). Par l'intermédiaire de plugin, il est possible d'insérer des visualisations graphiques proposant plusieurs vues de l'activité des membres du réseau social des veilleurs (ex : le graphe social égo-centré personne / personnes / tags partageant les mêmes concepts manipulés, une timeline des tags les plus cités sur une ligne de temps, etc.). Des modèles et des méthodes (Erétéo 2011) permettent de représenter les acteurs, leurs relations (FOAF, RELATIONSHIP), leurs activités en ligne (SIOC), et de structurer les concepts qu'ils manipulent (SKOS).

Plusieurs expériences ont déjà pu être effectuées ou sont en cours sur l'infrastructure proposée, avec des collectifs souvent pluridisciplinaires atteignant jusqu'à 20 veilleurs combinant plusieurs des services évoqués. Notre objectif est de continuer la mise à l'épreuve de cette boite à outils de veille collaborative, pour vérifier cette approche est de nature à faciliter l'activité collective de veille.

Références

BENEL, A., LEJEUNE, C., ZHOU, C., Éloge de l'hétérogénéité des structures d'analyse de textes. Document numérique, RSTI 13(2), 41-56. Hermès-Lavoisier, 2010.

BAUMARD, P. (1991). Stratégie et surveillance des environnements concurrentiels. Masson.

BUFFA M., HUSSON G., DELAFORGE, N.. SweetDeki : le wiki sémantique couteau suisse du réseau social ISICIL.EGC, volume RNTI-E-23 of Revue des Nouvelles Technologies de l'Information, page 387-398. Hermann-Éditions, (2012)

BUFFA M., DELAFORGE N., ERETEO G., GANDON F., GIBOIN A. AND LIMPENS F., "ISICIL: Semantics and Social Networks for Business Intelligence", conference SOFSEM 2013, 39th Int.Conference on Current Trends in Theory and Practice of Computer Science. January 26-31, 2013 Spindler Mlýn, Czech Republic.

BULINGE, F. (2006). Le cycle du renseignement : analyse critique d'un modèle empirique. Market Management , pp 36 à 52.

CAHIER J.-P.,BENEL,A. SALEMBIER, P.,(2013).Towards a "non-disposable" software infrastructure for participation. Interaction Design and Architecture(s) Journal (IxD&A) 18, 68–83. Univ. Roma II.

DELAFORGE N.,GANDON F.,Webmarks: Le marquage d'intérêt sur le Web de données, 12e Conf. Extraction & Gestion des Connaissances(EGC), Bordeaux, France, 2012

ERÉTÉO, G. (2011). Semantic social Network Analysis. PhD Thesis.

LEJEUNE CH., BENEL A., Lexicométrie pour l'analyse qualitative : Pourquoi et comment résoudre le paradoxe.Actes des 11e journées internationales d'analyse statistique de données textuelles (JADT), Lexicometrica.2012.

LEITZELMAN, M. (2010, janv). 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 , Volume 6, p.200.

WEICK, K. (1995). Sensemaking in organizations. Thousand Oaks: Sage Publications.

ZAHER L'H., CAHIER J.-P., ZACKLAD M., The Agoræ/Hypertopic approach, Proceedings of the workshop on Indexing and Knowledge in Human Sciences, Nantes, June 26-28, 2006. 66-70.. 2006.

ZHOU C., LEJEUNE CH. AND BÉNEL A., Towards a standard protocol for community-driven organizations of knowledge, in Proceedings of the 13th International Conference on Concurrent Engineering (ISPE CE'06), (2006) IOS Press.