



HAL
open science

EVARIST: un outil de monitoring du buzz et de l'e-reputation sur Twitter

Etienne Cuvelier, Marie-Aude Aaufaure

► **To cite this version:**

Etienne Cuvelier, Marie-Aude Aaufaure. EVARIST: un outil de monitoring du buzz et de l'e-reputation sur Twitter. 9e Atelier Visualisation et extraction de connaissances, 11ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances - EGC 2011, Brest, France, Du 25 au 28 janvier 2011., Jan 2011, Brest, France. pp.1-8. hal-00704365

HAL Id: hal-00704365

<https://hal.science/hal-00704365v1>

Submitted on 5 Jun 2012

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.

EVARIST: un outil de monitoring du buzz et de l'e-reputation sur Twitter

Etienne Cuvelier*, Marie-Aude Aufaure*

* Équipe Business Intelligence
Laboratoire Mathématiques Appliquées aux Systèmes
École Centrale Paris,
Etienne.Cuvelier@ecp.fr, Marie-Aude.Aufaure@ecp.fr

Résumé. Dans le monde interconnecté actuel, la vitesse de diffusion des informations amène à une formation des opinions tendant vers de plus en plus d'immédiateté. En effet, les grands réseaux sociaux, en permettant le partage, et donc, la diffusion de l'information de manière quasi-instantanée, accélèrent aussi la formation des opinions concernant l'actualité. Cela fait de ces réseaux de formidables observatoires des opinions et aussi de l'e-réputation. Dans cet article nous proposons un prototype d'outil, basé sur les treillis de Galois, nommé EVARIST, qui permet à un utilisateur, institutionnel ou individuel, à partir d'un ensemble de mots clés de son choix (thème, marque, nom propre,...) de visualiser l'ensemble des termes les plus associés circulant à travers le réseau Twitter, et formant donc l'actualité brûlante (buzz) concernant le sujet choisi.

1 Introduction

Depuis leur apparition, les blogs et les réseaux sociaux suscitent un intérêt croissant pour l'observation et la modélisation de l'opinion, comme le démontre, notamment, la session spéciale que leur consacrent les conférences TREC depuis l'édition de 2006 (Ounis et al. (2006)). La détection de sujets d'actualité brûlante fut ainsi une des tâches de l'édition 2009 de cette session blogs (voir Macdonald et al. (2009) pour une présentation). De même, les réseaux sociaux tels Facebook et Twitter, de par leurs fonctions de partage et de transfert de l'information devraient aussi permettre d'observer quasiment en temps réel la formation des opinions et, permettre ainsi de détecter les tendances. Par exemple Kramer (2010) utilise les mots exprimant les émotions dans les statuts Facebook des utilisateurs américains pour synthétiser un index modélisant le concept de "Bonheur National Brut".

Ces réseaux sociaux sont donc des lieux d'observation privilégiés pour de la formation des opinions, notamment à propos d'un sujet choisi, que cela soit une personne (personnal branding), une institution officielle ou un opérateur industriel. Dans le cas de l'e-reputation, l'observation du buzz et plus particulièrement du buzz négatif (bad buzz) concernant le sujet choisi est particulièrement importante. Nous avons testé cette observation sur le plus réactif de ces réseaux, à savoir, Twitter.

Cet article s'organise de la façon suivante : dans la section 2 nous détaillons succinctement comment fonctionne le réseau Twitter, quelles sont ses contraintes et ses conventions, et les

difficultés d'analyse qu'elles entraînent. Dans la section 3 nous introduisons l'analyse formelle de concepts et les treillis de Galois, qui permettent de contourner les difficultés évoquées. Enfin dans la section 4 nous exposons les principes base de notre outil EVARIST, ainsi que les résultats obtenus sur un petit ensemble d'information relatives au mot clé "e-réputation". Nous terminons avec les conclusions et perspectives de développement.

2 Twitter et Micro-Blogging

Twitter a été créé en 2006 dans le but de permettre à ses utilisateurs de partager facilement de courts messages textuels appelés *tweets*¹. Le système ayant été initialement conçu pour partager les tweets via SMS, une limite de 140 caractères a été fixée à ces messages courts. Bien que le système soit aujourd'hui massivement utilisé via le web et via des applications développées pour ordinateurs ou smartphones, cette contrainte de 140 caractères n'a jamais été levée. L'aspect réseau social de cette plateforme réside dans son aspect orienté, et ses principes de base sont les suivants :

- un utilisateur peut, avec son compte Twitter, générer ou transmettre de l'information via un champ de saisie ;
- un utilisateur A peut suivre les tweets d'un utilisateur B en le signalant via l'interface ad hoc, et ce, sans que B ne doive suivre les tweets de A en retour.

Les utilisateurs qui suivent un compte Twitter A s'appellent ses *abonnés* ou *followers*, alors que les utilisateurs que A suit s'appellent ses *abonnements* ou *followings*. L'ensemble des tweets des abonnements d'un compte s'appelle sa *timeline*.

De par son principe de "micro-publications", Twitter permet un partage et une diffusion très rapides de l'information ainsi que des opinions concernant cette dernière. La croissance de ce service est actuellement importante et, au mois d'avril 2010, Twitter comptait presque 6 millions d'utilisateurs enregistrés, 300 000 nouveaux comptes par jour et, en moyenne, 55 millions tweets générés par jour (voir Bosker (2010)). Cette intense activité se traduit par une très grande réactivité par rapport aux faits d'actualité, et cette réactivité se révèle extrêmement intéressante pour l'analyse de la formation d'opinions. Ainsi O'Connor et al. (2010) ont montré qu'il existait une corrélation très importante entre trois indices existants, calculés via des enquêtes quotidiennes, concernant la confiance des consommateurs américains, les sondages Obama/Mc Cain pendant la campagne présidentielle américaine et ensuite l'appréciation du travail du président Obama, et les opinions formulées sur Twitter à propos de ces sujets. De manière plus prédictive Ritterman et al. (2009) ont montré que l'information circulant sur Twitter concernant la grippe aviaire permettait, associée à un modèle de prédiction de marché, de prédire plus efficacement l'opinion concernant la transformation de la grippe en pandémie.

Un certain nombre de conventions ont cours sur Twitter, nous allons en relever maintenant les principales, nécessaires à la compréhension minimale de la plateforme. La première de ces conventions est l'utilisation de l'arobase pour citer ou s'adresser à un utilisateur précis, ainsi dans l'exemple ci-dessous, l'utilisateur Jules s'adresse à l'utilisateur Jim en citant l'utilisatrice Catherine :

Jules: @Jim rendez-vous chez @Catherine à 22h00?

1. Gazouillis en anglais.

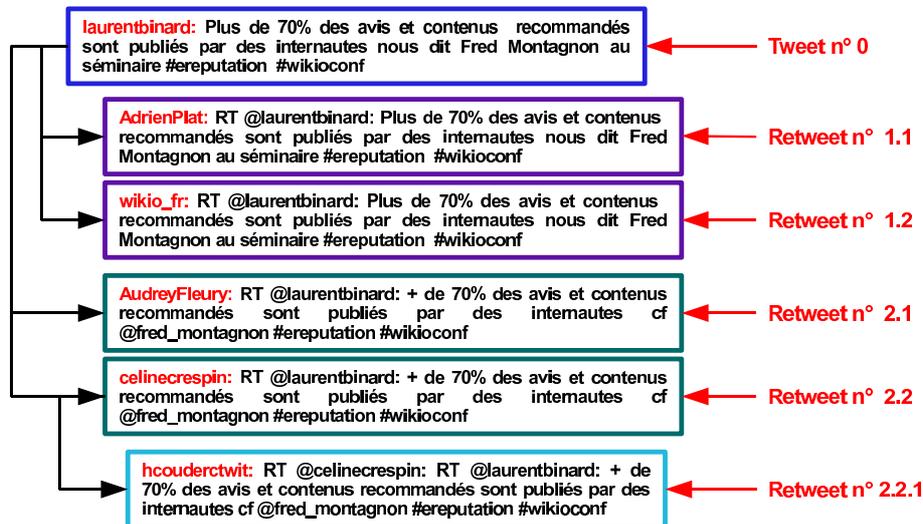


FIG. 1 – Illustration du polymorphisme de l'information relayée.

Les utilisateurs Jim et Catherine verront apparaître ce tweet dans leurs timelines respectives. Une seconde convention est la pratique du *retweet* ou *RT*. Lorsqu'un utilisateur voit dans sa timeline une information qu'il veut partager avec ses abonnés, il utilisera la fonction retweet du service (web ou application) qu'il utilise, comme illustré dans l'exemple ci-dessous :

Jules: Inception est génial.

Jim: Bof RT @Jules: Inception est génial.

Catherine: LOL RT @Jim: Bof RT @Jules: Inception est génial.

Bien que la majorité des retweets soient annoncés par "RT @", d'autres variantes et pratiques coexistent, comme cela a été très bien analysé dans Boyd et al. (2010). Par exemple certains utilisateurs éditent le retweet en ajoutant, par exemple un (*via @*) à la fin de celui-ci, comme dans l'exemple suivant :

Jules: <http://www.google.com> is awesome!

Catherine: RT @Jules: <http://www.google.com> is awesome!

Jim: <http://www.google.com> is awesome! (via @Jules)

Cette possibilité d'éditer un retweet, associée à la contrainte des 140 caractères crée ce que nous appellerons le *polymorphisme de l'information relayée* sur Twitter. Ce polymorphisme est illustré dans la figure 1. Dans cet exemple on voit que le tweet initial (tweet no 0) a été retweeté de plusieurs façons. On voit un premier groupe de retweets (no 1.1 et no. 1.2) dans lesquels le tweet initial est repris tel quel, et un second groupe de retweets (no 2.1 et no 2.2), où les utilisateurs ont modifié légèrement le retweet. De plus un des retweets de ce deuxième groupe est lui même retweeté tel quel (tweet no 2.2.1). Ce polymorphisme pose un réel problème lorsque l'on veut mesurer la popularité d'une information, car si l'on se limite à compter le nombre de retweets non-modifiés, on omettra de comptabiliser une partie des retweets, qui malgré leurs modifications véhiculent essentiellement la même information.

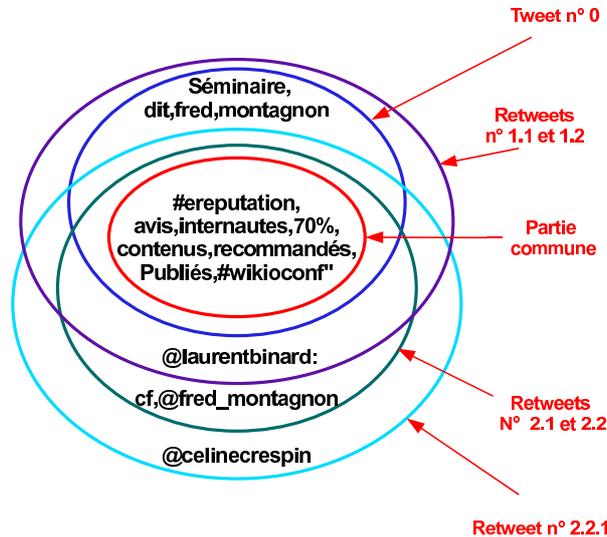


FIG. 2 – Vue ensembliste du polymorphisme de l'information relayée.

Néanmoins, si l'on fait abstraction des mots de liaisons et des signes de ponctuations (en l'occurrence dans notre exemple : plus, +, de, des, et, sont, par, nous), que l'on se restreint aux mots significatifs, alors on peut représenter l'information relayée sous forme ensembliste, comme l'illustre la figure 2. On voit dans cette représentation que les inclusions successives permettent de visualiser la partie commune à tous les tweets, le noyau de l'information relayée. L'inclusion définit aussi une relation d'ordre partiel qui peut être représentée par un diagramme de Hasse comme dans la figure 3. L'établissement d'un tel diagramme permet d'établir, à la fois, quels sont les mots communs à un ensemble de tweets, ainsi que les différentes formes sous lesquelles une même information a été relayée. L'analyse formelle de concepts et les treillis de Galois permettent exactement de structurer l'information sous cette forme.

3 Analyse Formelle de Concepts - Treillis de Galois

L'analyse formelle de concept, en anglais *Formal Concept Analysis*, *FCA*, (Wille (1980)) est basée sur les *Treillis de Galois* (Barbut et Monjardet (1970) et Birkhoff (1940)), qui peuvent être utilisés pour la classification conceptuelle (Carpineto et Romano (1993) et Wille (1984)).

Un treillis de Galois permet de regrouper, de façon exhaustive, des objets en classes, appelées *concepts*, en utilisant leur propriétés partagées. Un treillis est classiquement basé sur une matrice booléenne, appelée *matrice de contexte* et notée C , dont les lignes représentent un ensemble d'*objets* O que l'on souhaite décrire, et les colonnes, un ensemble d'*attributs* A que ces objets ont ou non et, qui permettent donc la description des dits objets. Nous utiliserons, pour introduire ces treillis, un exemple simple que l'on peut trouver dans Wolff (1993). Supposons que nous avons une description des espèces animales suivantes (voir table 1) : Lion, Moineau, Aigle, Lièvre, Autruche, Abeille et Chauve-souris. Description basée sur la liste des propriétés que les animaux de l'espèce possèdent ou non : Sont-ils prédateurs ? Volent-ils ? Sont-ce des

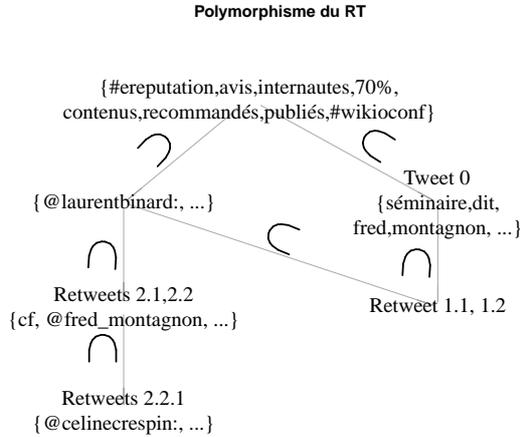


FIG. 3 – Analyse du polymorphisme du retweet de l’information.

	Prédation	Vol	Oiseau	Mamifère
Lion	x			x
Moineau		x	x	
Aigle	x	x	x	
Lièvre				x
Autruche			x	
Abeille		x		
Chauve-souris		x		x

TAB. 1 – Table de contexte pour les espèces animales.

oiseaux ? Sont-ce des mammifères ? La possession de la propriété $a \in A$ par l’objet $o \in O$ traduit l’existence d’une relation I entre eux : aIo . L’existence de cette relation I entre O et A est matérialisée dans la matrice de contexte C par, soit la valeur “vrai” (et “faux” sinon) soit par une marque quelconque (et rien sinon) . Le triplet $K = (O, A, I)$ est appelé un *contexte formel* ou simplement un contexte.

L’*intention* d’un ensemble $X \subset O$ est l’ensemble des attributs possédés conjointement par tous les objet de X et, est donnée par la fonction f :

$$f(X) = \{a \in A | \forall o \in X, oIa\}. \tag{1}$$

Inversement l’*extension* d’un ensemble $Y \subset A$ est l’ensemble des objets à posséder conjointement tous les attributs de Y et, est donnée par la fonction g :

$$g(Y) = \{o \in O | \forall a \in Y, oIa\}. \tag{2}$$

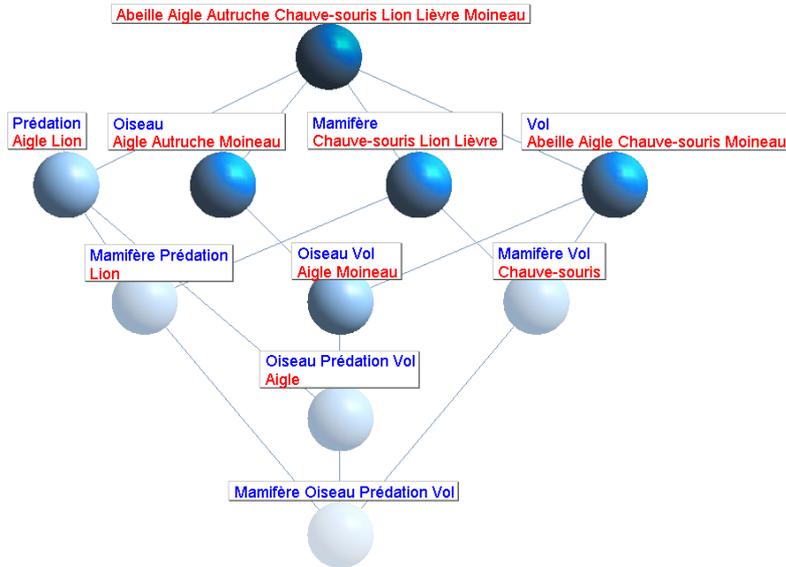


FIG. 4 – Affichage plein du treillis de Galois des espèces animales de la table 1.

Le couple (f, g) est appelé une *connexion de Galois*.

Un *concept* est tout couple $C = (X, Y) \subset O \times A$, tel que les objets de X soient les seuls à posséder tous les attributs de Y , en d’autres termes $X \times Y$ forme, à deux permutations près de O et de A , un rectangle maximal dans C , c’est-à-dire que

$$f(X) = Y \ \& \ g(Y) = X. \quad (3)$$

Pour illustrer cette notion de concept on peut observer dans la table 1 que l’ensemble $X = \{\text{Moineau}, \text{Aigle}\}$ donne un concept car $f(X) = \{\text{Vol}, \text{Oiseau}\} = Y$ et $g(Y) = X$, et ce concept est donc $(\{\text{Moineau}, \text{Aigle}\}, \{\text{Vol}, \text{Oiseau}\})$, alors que l’ensemble $X' = \{\text{Lion}, \text{Lièvre}\}$ ne donne pas un concept car $f(X') = \{\text{Mammifère}\} = Y'$ and $g(Y') = \{\text{Lion}, \text{Lièvre}, \text{Chauve-souris}\}$, mais par contre ce dernier ensemble, lui, donne le concept $\{\text{Lion}, \text{Lièvre}, \text{Chauve-souris}\}, \{\text{Mammifère}\}$.

Le *treillis de Galois* est le *poset* de concepts L muni de l’ordre partiel suivant \leq :

$$(X_1, Y_1) \leq (X_2, Y_2) \Leftrightarrow X_1 \subseteq X_2 \text{ (or } Y_1 \supseteq Y_2). \quad (4)$$

Le treillis de Galois est noté $T = (L, \leq)$ et, est représenté à l’aide d’un *diagramme de Hasse* comme dans les figures 4 et 5 pour les espèces. Deux types d’affichage existent pour les labels des concepts, l’affichage plein et l’affichage réduit. Dans l’affichage plein, tous les objets et attributs d’un concept sont affichés, comme dans la figure 4. Dans l’affichage réduit, les attributs et les objets ne sont affichés qu’une seule fois, la première fois qu’ils sont rencontrés en parcourant le treillis, en partant du haut pour les attributs, et en partant du bas pour les objets, comme c’est le cas dans la figure 5.

Le calcul du treillis peut être effectué à l’aide, par exemple, de l’algorithme de Bordat (1986), qui calcule récursivement tous les concepts à partir du concept $(\emptyset, f(\emptyset))$, en calculant

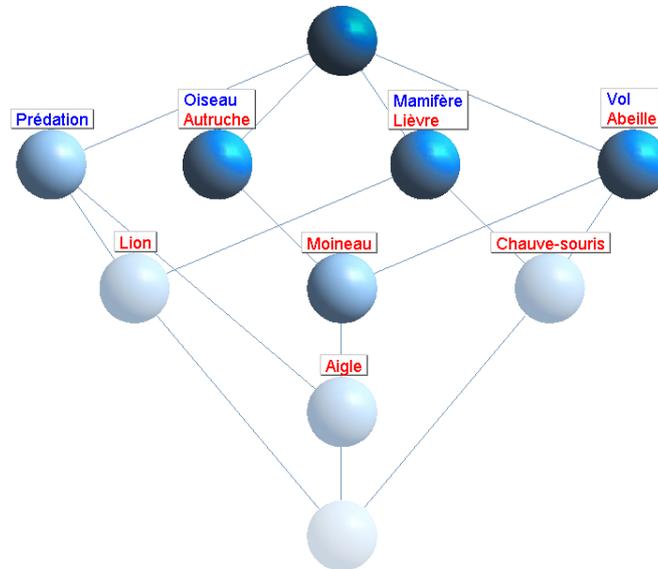


FIG. 5 – Affichage réduit du treillis de Galois des espèces animales de la table 1.

pour chaque concept trouvé l'ensemble de ses concepts-fils. Pour voir une revue des autres algorithmes pour la génération de treillis, voir l'article de Kuznetsov et Obedkov (2001) qui effectue aussi une comparaison des performances.

Un des avantages non négligeable de la classification basée sur les treillis est que pour une table de contexte donnée le treillis résultant est unique (pas d'instabilité à l'exécution), et il est exhaustif (tous les concepts existants s'y trouvent). Dans le cas qui nous occupe cette classification va nous permettre de retrouver tous les groupes de mots présents dans un groupe de tweets et de les représenter de façon similaire à la figure 3.

4 E-buzz Monitoring

Dans ce qui suit nous proposons d'analyser un groupe de tweets, afin d'y repérer les termes et groupes de termes les plus tweetés dans le groupe considéré. Pour ce faire nous proposons une démarche en quatre étapes :

1. Récupération des tweets contenant un ou plusieurs mots clés ;
2. Nettoyage des tweets (suppression des mots de liaisons, ponctuations,...) ;
3. Établissement d'une table de contexte où les objets sont les tweets et, les attributs les mots contenus dans ces derniers ;
4. Calcul du treillis de Galois correspondant ;
5. Visualisation des résultats.

Pour illustrer notre technique nous utiliserons un ensemble de 50 tweets récupérés en faisant une recherche sur le mot clé "#ereputation". Nous donnons ci dessous, en guise d'illustration les 5 premiers tweets de cet ensemble :

EVARIST: buzz et e-reputation monitoring

Tweet 1: overclub: #ereputation : votre avis sur les multiples solutions de veille... d'après vous quel est l'outil le plus efficace ?
Tweet 2: AudreyFleury: #eReputation Les internautes veillent sur leur eRéputation Stratégies <http://ow.ly/1a7fWM>
Tweet 3: AudreyFleury: RT @laurentbinard: + de 70% des avis et contenus recommandés sont publiés par des internautes cf @fred_montagnon #ereputation #wikioconf
Tweet 4: hcouderctwit: RT @celinecrespin: RT @laurentbinard: + de 70% des avis et contenus recommandés sont publiés par des internautes cf @fred_montagnon #ereputation #wikioconf
Tweet 5: wikio_fr: RT @laurentbinard: Au séminaire #wikioconf, Serge Alleyne, fondateur de #nomao, annonce et présente sa solution de #ereputation locale avec #wikiobuzz...

L'application des étapes 1 à 4 ne pose pas de problème sur un treillis de taille modeste, alors que la cinquième, la visualisation des résultats, elle, est moins évidente. Dans la figure 6 nous affichons l'ensemble du treillis en donnant à chaque concept une surface qui soit proportionnelle au nombre de tweets qu'il contient. On constate que malgré sa petite taille (59 concepts trouvés) il est difficile d'afficher tout les concepts proportionnellement à leur taille et, simultanément, l'ensemble de leurs attributs de façon lisible, et ce même en utilisant l'affichage réduit des attributs. Pour réduire le nombre de concepts à afficher, nous pouvons sélectionner que les concepts dont la taille relative (nombre d'objets du concept divisé par le nombre d'objets dans la table de contexte) dépasse un seuil choisi, ce qui est cohérent avec le concept de buzz, car ce sont les groupes de mots les plus tweetés. C'est ce qui est fait dans la figure 7 avec un seuil de 10%, mais cela n'augmente pas suffisamment la lisibilité, car les attributs de notre treillis sont formés de mots ou groupes de mots plus ou moins longs. Un autre type de visualisation s'avère donc nécessaire pour ce groupe de concepts avec les mots les plus retweetés. On peut bien entendu essayer d'afficher les mots contenus dans les différents concepts à l'aide de nuage de tags, en donnant aux mots du concept une taille proportionnelle à l'importance de celui-ci. Néanmoins, même s'il existe des solutions pour afficher les tags associés l'un près de l'autre comme dans Kaser et Lemire (2007), on perd avec cet affichage les liaisons d'inclusion entre un concept et ses sous-concepts. De plus un sous-concept pouvant dépendre de plusieurs super-concepts, cela risque de complexifier la tâche de regroupement des tags proches. Nous proposons donc d'afficher les concepts les plus importants sous forme de nuage de tags, proportionnels et en réseau, c'est-à-dire que les liaisons entre les différents concepts seront matérialisées par des arrêtes. Ces arrêtes seront dirigées, allant du concept vers ses sous-concepts. Nous effectuons le placement des noeuds à l'aide de la méthode Fruchterman et Reingold (1991), car cette méthode optimise l'écartement des noeuds et permet ainsi d'augmenter la lisibilité des tags. Enfin, pour renforcer une lecture allant du plus général au plus particulier, nous avons décidé d'ajouter une allégorie topographique, similairement aux cartes topographiques proposées dans Fujimura et al. (2008). Pour ce faire, à chaque point du graphique résultant, nous ajoutons un niveau, les différents niveaux étant représentés à l'aide des classiques courbes de niveaux. Pour la mise aux points des niveaux, nous utilisons un mélange de gaussiennes à deux dimensions transformées, auxquelles nous donnons comme centres les coordonnées des centres des tags, et comme écart-types les largeurs et hauteurs des dits tags. Enfin, pour donner des hauteurs qui soient proportionnelles aux tailles des concepts exactement aux centres des tags, nous normalisons les hauteurs des gaussiennes en les multipliant par les

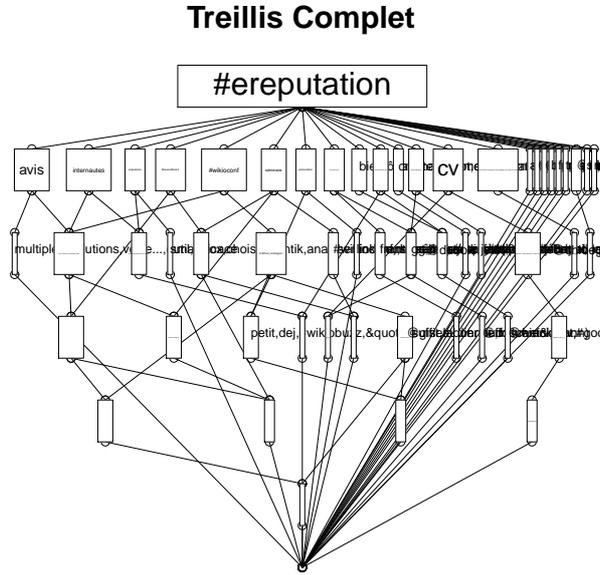


FIG. 6 – *Le treillis de Galois des tweets.*

écart-types et ensuite par hauteurs désirées. Le mélange résultant est la fonction topographique T :

$$T(x, y) = \sum_{i=1}^k \frac{s_i}{2\pi} e^{-\frac{(x-x_i)^2 (x-y_i)^2}{2l_i^2 h_i^2}} \quad (5)$$

où :

- k représente le nombre de concepts affichés,
- x_i et y_i représentent les coordonnées du i^e concept,
- l_i et h_i représentent la largeur et la hauteur du tag du i^e concept,
- s_i représente la taille du i^e concept.

Bien entendu, comme nous avons changé les volumes sous les surfaces, notre fonction topographique T n'est plus une densité de probabilité, mais cette propriété ne nous est pas nécessaire ici.

Le résultat final peut être vu en figure 8. Dans cette figure, à partir du concept représentant le mot clé de départ $\{#ereputation\}$, on voit que les sous-concepts les plus importants sont $\{avis\}$, $\{internaudes\}$ et $\{#wikiconf\}$, et que ces trois concepts contiennent aussi le concept $\{70\%, contenus, recommandés, publiés\}$, alors que seul le concept $\{#wikiconf\}$ contient le concept $\{cf, @fred_montagnon\}$. D'autre part on voit trois concepts affichés indépendamment des premiers : $\{c'est\}$, $\{@laurentbinard : \}$ et $\{daily, read, twitter, newspaper, http ://bit.ly/d, (19, contributions, todays)\}$. Et enfin le concept $\{cv\}$ qui dépend du concept $\{c'est\}$. L'idée de cette visualisation est de laisser glisser le regard du lecteur des “sommets” (les concepts les

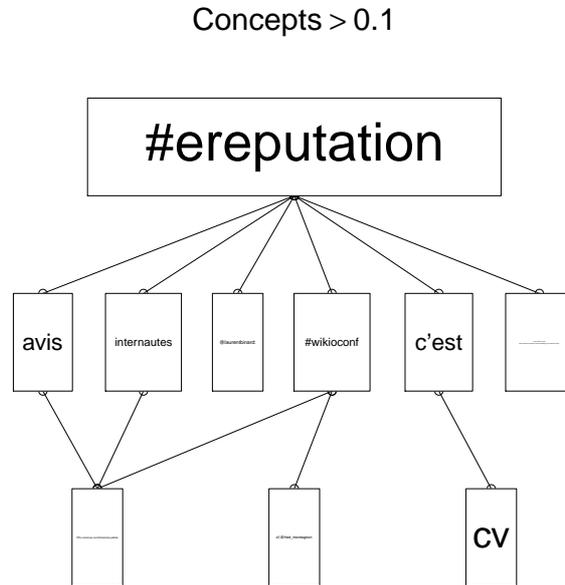


FIG. 7 – Les concepts comprenant plus de 10% des tweets.

plus généraux) vers les “vallées” (les concepts les plus particuliers). La construction et les affichages des treillis et des cartes de tags en réseau topographique ont été effectués dans l’environnement statistique R Development Core Team (2010), et pour les parties treillis, à l’aide du package *galois* développé par nos soins.

5 Conclusions et perspectives

Dans cet article nous avons présenté une nouvelle technique de monitoring du buzz et de la e-réputation sur la plateforme de micro-blogging, Twitter. Cette technique se base sur les treillis de Galois, et propose comme visualisation des concepts résultants, un nuage de tags proportionnels en réseau topographique. Cet affichage, limité aux concepts les plus importants, permet d’afficher les tags constituant les concepts de manière plus lisible que si l’on affiche directement le treillis. Son idée est de faire “glisser” le regard du lecteur, des concepts les plus généraux, affichés aux “sommets” vers les concepts plus particuliers placés dans les “vallées”, les flèches du réseaux servant de “pistes” pour guider vers les concepts liés.

Bien qu’EVARIST soit encore à l’état de prototype, un certain nombre d’améliorations sont envisageables et envisagées. La première c’est l’ajout de l’interactivité afin de permettre à l’utilisateur de sélectionner le sous-concept qu’il désire développer. On envisage ensuite de “développer” les URL réduites (bit.ly, is.gd, tinyURL,..) afin de ne pas comptabiliser sous deux concepts différents une même URL qui aurait été réduite via deux services différents. Lors de l’étape qui consiste en la suppression des mots de liaisons et des signes de ponctuations afin

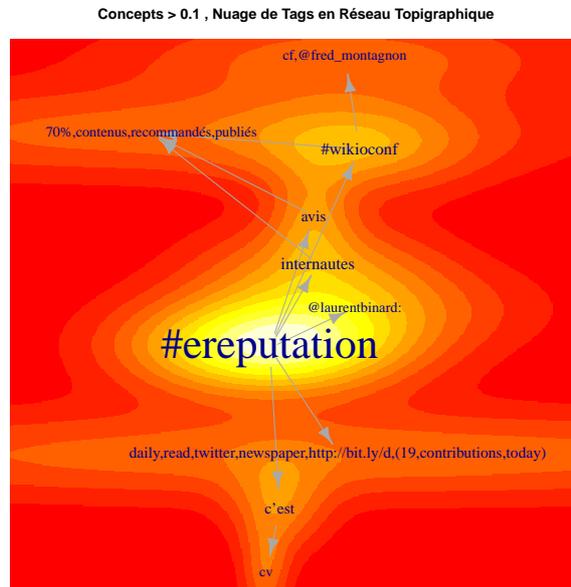


FIG. 8 – Le réseau topographique des concepts comprenant plus de 10% des tweets.

d'établir la table de contexte, un traitement particulier devrait être réservé aux smileys, qui sont en fait signifiants. Dans une prochaine étape, nous prévoyons aussi d'utiliser l'analyse de sentiments pour évaluer la positivité ou négativité des concepts, notions importantes dans le cadre de la surveillance de la réputation électronique. Enfin le support de plusieurs langues se révèle un challenge important mais intéressant pour le développement d'un tel outil.

Références

- Barbut, M. et B. Monjardet (1970). *Ordre et classification, Algèbre et combinatoire, Tome 2*. Hachette.
- Birkhoff, G. (1940). *Lattice Theory*, Volume 25. New York : American Mathematical Society.
- Bordat, J. (1986). Calcul pratique du treillis de galois d'une correspondance. *Mathématique, Informatique et Sciences Humaines* 24(94), 31–47.
- Bosker, B. (2010). Twitter user statistics revealed. <http://www.huffingtonpost.com/>.
- Boyd, D., S. Golder, et G. Lotan (2010). Tweet, tweet, retweet : Conversational aspects of retweeting on twitter. In *Proceedings of the 43rd Hawaii International Conference on Social Systems (HICSS)*.
- Carpineto, C. et G. Romano (1993). Galois : An order-theoretic approach to conceptual clustering. In *Proc. Of the 10th Conference on Machine Learning, Amherst, MA, Kaufmann*, pp. pp. 33–40.

EVARIST: buzz et e-reputation monitoring

- Fruchterman, T. et E. Reingold (1991). Graph drawing by force-directed placement. *Software - Practice and Experience* 21(11), 1129–1164.
- Fujimura, K., S. Fujimura, T. Matsubayashi, T. Yamada, et H. Okuda (2008). Topigraphy : visualization for large-scale tag clouds. In *Proceeding of the 17th international conference on World Wide Web*, WWW '08, New York, NY, USA, pp. 1087–1088. ACM.
- Kaser, O. et D. Lemire (2007). Tag-cloud drawing : Algorithms for cloud visualization. In *WWW2007 Workshop on Tagging and Metadata for Social Information Organization*, Banff, Alberta.
- Kramer, A. D. I. (2010). An unobtrusive behavioral model of gross national happiness. In *Proceedings of the 2010 conference on Human Factors and Computing Systems (CHI 2010)*.
- Kuznetsov, S. O. et S. A. Obedkov (2001). Comparing performance of algorithms for generating concept lattices. In *Concept Lattices-based Theory, Methods and Tools for Knowledge Discovery in Databases (CLKDD'01)*, Stanford, July 30, 2001.
- Macdonald, C., I. Ounis, et I. Soboroff (2009). Overview of the trec 2009 blog track. In *NIST Special Publication 500-278 : The Eighteenth Text REtrieval Conference Proceedings (TREC 2009)*.
- O'Connor, B., R. Balasubramanyan, B. R. Routledge, et N. A. Smith (2010). From tweets to polls : Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC*.
- Ounis, I., C. M. M. de Rijke, G. Mishne, et I. Soboroff (2006). Overview of the trec 2006 blog track. In *NIST Special Publication 500-272 : The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, Volume 272, pp. 17–31.
- R Development Core Team (2010). *R : A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Ritterman, J., M. Osborne, et E. Klein (2009). Using prediction markets and twitter to predict a swine flu pandemic. In *1st International Workshop on Mining Social Media*.
- Wille, R. (1980). *Restructuring lattice theory, Ordered sets I*. Rival.
- Wille, R. (1984). Line diagrams of hierarchical concept systems. *Int. Classif.* 11, 77–86.
- Wolff, K. E. (1993). A first course in formal concept analysis - how to understand line diagrams. In F. Faulbaum (Ed.), *SoftStatt'93, Advances in Statistical Software* 4, 429-438.

Summary

In the actual interconnected world, the speed of broadcasting of information leads the formation of opinions towards more and more immediacy. Big social networks, by allowing distribution, and therefore broadcasting of information in a almost instantaneous way, also speed up the formation of opinions concerning actuality. Then, these networks are great observatories of opinions and e-reputation. In this article we propose a prototype of tool, based on Galois lattices, which allows an user, institutional or individual, from a chosen set of key words (topic, mark, proper name) to show all the most linked terms circulating across Twitter, and forming therefore the hot topics (Buzz) concerning the chosen subject.