Analyse d'activité et exposition de la vie privée sur les médias sociaux

Younes Abid*, Abdessamad Imine*, Amedeo Napoli*,
Chedy Raïssi*, Marc Rigolot**, Michaël Rusinowitch *

*INRIA-Nancy, 54600 Villers-lès-Nancy
prenom.nom@inria.fr
** Fondation MAIF, 50 avenue Salvador Allende , 79000 Niort
prenom.nom@fondation.maif.fr

Dans ce travail [1] nous avons réalisé une enquête sur l'usage des médias sociaux pour déterminer les sujets sensibles et détecter des vulnérabilités de vie privée. Nous avons collecté 232 réponses complètes et valides d'utilisateurs de médias sociaux. La corrélation par rapport à la variable "âge" entre notre échantillon et la population des internautes français [2] est 0,8 et s'élève à 0,95 pour les internautes de plus de 18 ans. Nous avons analysé le comportement des internautes sur les médias sociaux suivant quatre critères et défini les sujets sensibles comme étant ceux qui appartiennent à *au moins deux* ensembles parmi les suivants : L'ensemble $E_{discussion}$ des sujets dont la fréquence de discussion globale est inférieure à la fréquence moyenne moins l'écart type. Dans notre étude $E_{discussion}$ est {"*Argent*", "*Achats*", "*Religion*", "*Rencontre*"}. L'ensemble $E_{activite}$ des forums et sites internet dont le taux d'activité globale est inférieur au taux moyen moins l'écart type est {"*Sortie, Rencontre, Chat*", " *Philosophie, Religion, Libre pensée*"}. L'ensemble $E_{anonyme}$ des sites et forums dont le taux de publication anonyme (sans identification ou avec des profils anonymes) dépasse la moyenne de 8.7 % est {"*Économie, Politique, Actualité, Infos*", "*Philosophie, Religion, Libre pensée*", "*Jeux, Musique, Film, Humour, Art, Livre*", "*Santé, Courses, Cuisine, Maison, Astuce*"}. Nous avons simulé des entretiens individuels directifs pour identifier les sujets évités sur les réseaux sociaux. Les participants avaient la possibilité de développer une réponse libre dans sa forme et dans sa longueur. Nous avons ensuite analysé les réponses et défini l'ensemble E_{evite} des sujets évités en se basant sur les sujets mentionnés et les mots répétés fréquemment par les répondants. E_{evite} = {"*Politique*", "*Religion*", "*Vie personnelle et familiale*" , "*Vie sentimentale et sexuelle*", "*Vie financière*", "*Actualité*", "*Vie professionnelle*", "*Santé*", "*Art*" "*Vacances et Voyages*" }

Nous avons normalisé les séries de pourcentages calculées dans notre enquête pour les rendre comparables. La transformation des données a consisté à diviser chaque valeur par la moyenne de la série. Étant donné un sujet : moins il est discuté sur des médias sociaux plus il est sensible. Aussi, nous définissons le *coefficient de sensibilité* C par opposition au taux de discussion sur les médias sociaux. Le tableau 1 classe les sujets et les données personnelles inférées du plus sensible vers le moins sensible sur les médias sociaux.

1. réalisé dans le cadre d'un projet financé par la Fondation MAIF.

2. http://vingthuitzerotrois.fr/marketing/attachment/repartition-age-reseaux-sociaux/