

Sujet de thèse en CIFRE

Qualité de données dans la e-réputation

Le vivier de nouvelles technologies existant et qui ne cesse de croître permet aux utilisateurs de créer facilement du contenu et de l'annoter, créant une dynamique du Web avec un espace avec plus de consommateurs que de producteurs. Le Web est devenu une plate-forme où chaque consommateur de l'information est potentiellement un producteur. Par conséquent, Internet s'impose comme vecteur incontournable non seulement pour l'achat mais aussi pour la prescription. Ainsi, chaque marque souhaite soigner son image (subjective) c-a-d sa e-réputation qui peut être renvoyée à un internaute sur l'ensemble des contenus qui lui sont accessibles en ligne. Cette image peut être créée par les internautes ou par la marque elle-même, sur différentes sources : blogs, forums, réseaux sociaux, sites d'avis de consommateurs.

L'e-réputation est devenue de nos jours un des critères les plus importants pour la décision et le challenge pour les entreprises est de maîtriser leurs communications et leurs présences sur Internet pour mieux contrôler leur image. L'Internet est un grand terrain de jeu où une entreprise/marque peut tenter d'arbitrer les échanges d'informations. La difficulté réside dans l'absence de règles dans le web alors que l'internaute y est l'acteur principal : Il navigue quand et où il veut, il s'exprime comme il l'entend, il discute avec qui bon lui semble et il fait surtout plus de confiance à l'information qui vient des communautés d'utilisateurs et de la marque. Ces aspects affectent grandement la qualité des données sur un produit. Aussi, Une mauvaise qualité des données en entreprise peut avoir des conséquences très négatives sur son activité économique. Elle peut remettre en cause la production des indicateurs de performance et renvoie de fausses informations sur les résultats. La qualité des données est un objectif prioritaire pour toute entreprise souhaitant doper son efficacité, réduire ses coûts et améliorer sa relation client et améliorer sa e-réputation.

Par ailleurs, l'explosion des technologies autour des Big Data en entreprise offre la possibilité de faire face à la multiplication des sources d'informations, tels que les réseaux sociaux, systèmes de géo-localisation, données personnelles, web logs... afin d'accroître le niveau de connaissance de l'entreprise des comportements clients ou de l'usage de leur produits.

Le sujet de thèse s'inscrit dans la perspective ouverte pour le développement de nouvelles approches et méthodes de gestion de la qualité des données récoltées pour la e-réputation d'une entreprise à travers les réseaux sociaux. Autrement dit, Il s'agira de proposer un service aux entreprises leur permettant d'anticiper des actions à venir ou de les aider à prendre une décision précise relative à leur réputation tout en prenant en compte la qualité des données récoltées à partir des réseaux sociaux.

Mots clés : qualité de données, incohérences, incertitude, logique, inférence, résolution d'entité, machine learning.

Contacts

Salima Benbernou (salima.benbernou@parisdescartes.fr)

Mourad Ouziri (mourad.ouziri@parisdescartes.fr)

Lieu

Université Paris Descartes et la société IMBA consulting