



Programme Technologies logicielles 2006

Analyse d'image et de tendance sur les blogs

Février 2007 – Juillet 2009

Chef de file LINNA (Laboratoire d'Informatique de Nantes Atlantique)

Partenaires OverBlog, Sinequa, Syllabs (prestataire)

Résumé Les blogs sont aujourd'hui au coeur de l'actualité : ils prennent une importance de plus en plus grande. Lus par une population de bloggers qui représentent assez bien la population globale, ils couvrent toute l'étendue de la production de l'information. Il s'agit d'un nouveau pouvoir informationnel capable d'influencer les opinions des gens. Le but de ce projet est de développer des *outils de surveillance des blogs* qui permettent d'effectuer, automatiquement, deux tâches. La première est l'*étude d'image*. Le besoin sous-jacent est de créer une photographie de ce que pense le public d'une certaine personne, d'un organisme, etc. à un moment donné. La seconde tâche est l'*étude de tendance* dont le but est de déterminer, par exemple, les sujets émergents, l'état d'humeur d'une certaine population, l'évolution des sentiments vis à vis d'une certaine personne, d'un organisme, etc.

L'enjeu économique est évident : les personnalités et les entreprises se soucient de leur image. Elles doivent donc se tenir au courant de l'évolution de celle-ci ainsi que des sujets qui intéressent la population qu'ils ciblent, de façon à s'adapter à ses attentes.

Scientifiquement et techniquement, il existe principalement deux verrous :

1. Afin de pouvoir déterminer une image ou des tendances sur les blogs, il est nécessaire de disposer d'outils capables de considérer les sentiments. Or, la *langue des sentiments* est extrêmement complexe et une analyse ne peut se contenter de champs lexicaux. Les outils développés analyseront le texte dans un premier temps du point de vue axiologique (« succès » a une valeur axiologique positive alors que pour « échec » elle est négative) et dans un deuxième temps selon la modalité du sentiment invoquée (colère, haine, amour, etc.). Il s'agit là d'un axe de recherche de premier ordre et qui a des applications très importantes (dans le domaine de la veille en particulier).
2. Un autre point difficile est l'évaluation du *pouvoir d'influence* car il faut donner plus de poids à un blog qui a un pouvoir d'influence plus fort que les autres. Or, la détermination de ce pouvoir d'influence est difficile. Des critères subjectifs comme le « sérieux » du site entrent en jeu et sont beaucoup plus délicats à évaluer. Les blogs ne sont pas seulement représentatifs de l'opinion de l'auteur, ils sont aussi susceptibles

d'influencer les lecteurs. Il convient donc, pour faire l'analyse de l'image, d'évaluer le pouvoir d'influence dont dispose un blog. Il existe des indicateurs assez simples pour évaluer cela comme le nombre de visiteurs ou le nombre de liens pointant vers le blog en question. D'autres critères entrent en jeu comme le « sérieux » qui est lui beaucoup plus difficile à évaluer. Différents indicateurs statistiques et linguistiques seront considérés.