

# Analyse OLAP sur des tweets et des blogs : un retour d'expérience

Brice Olivier, Cécile Favre, Sabine Loudcher

Laboratoire ERIC, Université de Lyon, 5 Avenue Pierre Mendès-France  
69676 Bron Cedex, FRANCE  
{brice.olivier, cecile.favre, sabine.loudcher}@univ-lyon2.fr

**Résumé.** Le projet ANR IMAGIWEB dans lequel s'inscrit ce travail s'est donné pour mission d'étudier les images véhiculées sur Internet en se basant sur la détection d'opinions. Deux cas d'étude ont été définis : (1) le premier vise à répondre aux besoins d'analyse de chercheurs en science politique grâce à des données issues de Twitter durant la campagne présidentielle de 2012 ; (2) le second doit permettre à l'entreprise française EDF d'évaluer l'opinion du public en matière de sécurité, d'emploi et de prix à partir de billets de blogs. Dans cet article, nous présentons un retour d'expérience sur l'usage de l'analyse en ligne OLAP (OnLine Analytical Processing) pour des données textuelles, mettant en avant l'intérêt de ce type d'analyse pour les membres du projet.

## 1 Introduction

Le projet ANR IMAGIWEB<sup>1</sup> consiste à analyser et à suivre l'évolution de l'image (au sens de l'opinion) sur la toile, d'une part des personnages politiques à travers le réseau social Twitter, et d'autre part de l'entreprise EDF vis-à-vis du nucléaire en utilisant des blogs comme données. Ce projet regroupe différents partenaires parmi lesquels un laboratoire de recherche en science politique, des entreprises et des laboratoires de recherche en fouille de données.

Dans un premier temps, les tweets et blogs récoltés sont annotés manuellement pour relater l'opinion qu'ils véhiculent. Par la suite, l'enjeu sera de détecter automatiquement les opinions grâce à des méthodes de fouille d'opinion. Au-delà de la détection des opinions, pour mieux comprendre et analyser le contenu des tweets et des blogs, l'enjeu est aussi de les visualiser et de les explorer. Ainsi, un autre objectif du projet consiste à fournir à l'utilisateur, qu'il soit politologue, sociologue, marketeur ou encore analyste, un outil pour explorer les données (issues de tweets ou de blogs) et pour analyser en ligne l'opinion selon différents points de vue (sujets, temps, ...). L'analyse OLAP (OnLine Analytical Processing) permet de répondre à cet objectif de navigation, d'analyse et de visualisation.

L'OLAP sur des données textuelles correspond à une thématique de recherche récente avec des enjeux scientifiques importants. En effet, si l'OLAP a su montrer tout son potentiel analytique sur des "données classiques", la prise en compte de données textuelles nécessite une adaptation ou une évolution de l'OLAP pour prendre en compte les spécificités de ces données

---

1. ANR 2012-CORD- 002-01