

Analyse de gazouillis en ligne

Sandra Bringay ^{*,**}, Nicolas Béchet ^{***}, Flavien Bouillot ^{*}, Pascal Poncelet ^{*},
Mathieu Roche ^{*}, Maguelonne Teisseire ^{****}

^{*}LIRMM – CNRS, 161 rue Ada, 34095 Montpellier, France
{bringay,bouillot,poncelet,mroche}@lirmm.fr

^{**}Univ. Montpellier 3

^{***} INRIA Rocquencourt- Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France
nicolas.bechet@inria.fr

^{****} Cemagref, UMR TETIS, 500 rue Jean-François Breton, 34093 Montpellier, France
maguelonne.teisseire@cemagref.fr

Résumé. Les tweets échangés sur Internet constituent une source d'information importante même si leurs caractéristiques les rendent difficiles à analyser (140 caractères au maximum, notations abrégées, . . .). Dans cet article, nous définissons un modèle d'entrepôt de données permettant de valoriser et d'analyser de gros volumes de tweets en proposant des mesures pertinentes dans un contexte de découverte de connaissances. L'utilisation des entrepôts de données comme outil de stockage et d'analyse de documents textuels n'est pas nouvelle mais les mesures ne sont pas adaptées aux spécificités des données manipulées. Les résultats des expérimentations sur des données réelles soulignent la pertinence de notre proposition.

1 Introduction

Ces dernières années, le développement du web social et collaboratif 2.0 a rendu les internautes plus actifs au sein des réseaux participatifs. Les blogs pour diffuser son journal intime de manière massive, les dépêches RSS pour suivre les dernières informations sur un thème, les tweets pour publier ses faits et gestes, sont désormais extrêmement répandus. Simples à créer et gérer, ces outils sont utilisés par les internautes, les entreprises ou autres organisations pour communiquer à propos d'eux-mêmes ou des phénomènes qui les concernent. L'exploitation de ces nouvelles formes de publications s'inscrit dans une logique d'intelligence collective et suscite des applications inattendues en terme d'aide à la décision. En effet, les décideurs peuvent utiliser ces gros volumes d'information comme nouvelles ressources documentaires pour y puiser automatiquement de l'information.

Depuis son apparition en 2006, le site Twitter¹ s'est développé de telle manière qu'il est actuellement le 10^{ième} site le plus visité au monde². Twitter est une plate-forme de "micro-blogging". Il s'agit d'un système de partage d'informations via lequel les utilisateurs ont la

1. <http://twitter.com>

2. <http://www.alexa.com/siteinfo/twitter.com>