Projet ANR-2012-CORD-002-01

ImagiWeb

Programme CONTINT 2011

Résumé public (anglais)

Acronyme du projet	ImagiWeb
Titre du projet	Images sur le Web : Analyse de la dynamique des images sur le Web 2.0
Coordinateur du projet (société/organisme)	Laboratoire ERIC, Université de Lyon
Période du projet (date de début – date de fin)	1er Avril 2012 – 30 Septembre 2015
Site web du projet, le cas échéant	http://mediamining.univ-lyon2.fr/velcin/imagiweb

Rédacteur de ce rapport	
Civilité, prénom, nom	M. Julien VELCIN
Téléphone	04 78 77 24 14
Adresse électronique	Julien.Velcin@univ-lyon2.fr
Date de rédaction	27 Novembre 2015

Liste des partenaires présents à la fin du projet (société/organisme et responsable scientifique)	- AMI Software (E. Fourboul) - CEPEL, Université de Montpellier (J.Y. Dormagen) - EDF France (A. Peradotto) - ERIC, Université de Lyon (J. Velcin) - LIA, SFR Agorantic, Université d'Avignon et des Pays du Vaucluse (M. El-bèze) - Xerox Research Center Europe (C. Brun)

ImagiWeb: capturing online opinion with an approach combining both computer technique and sociological models

Objectives: We want to capture the opinion that has been expressed about an entity through combining computer science algorithms with a sociological paneling logic

Entities image (for instance celebrities, companies or brands) often comes to use through their virtual existence on the Web or on the other modern media. The goal of ImagiWeb is to analyze the opinion that was expressed in the messages that have been posted on the Internet about these entities with the help of statistical and computer science techniques. We want to study how these opinions relate to the social background of the individuals that did produce these messages through a novel paneling approach. To this resolutely multi-disciplinary approach, we want to add the possibility to assess these opinions through a set of specific targets that describe these entities (for instance: the price policy of a company or the way a politician is supported) and help follow the evolution of this image dynamically. This kind of research can have a technical impact (such as develop new algorithms and software), a strategic impact (bring new solutions to only reputation management) as well as a societal impact (to better understand the birth and broadcast of representation).

Methods and technologies: To combine a fine-grained analysis of extracted opinions from social media with an automatic category classification and paneling techniques

To solve this problem, we choose to combine advance Natural Language Processing tools (linguistic approach), text mining (statistical approach) and Machine Learning (supervised and unsupervised). The targets and polarities are yielded thanks to a hybrid and active method of supervised classification, whereas the gathering of homogeneous opinion groups is computed with an evolutionary probabilistic clustering. The message producers are identified through an original strategy of paneling of the users, thanks to a modernized version of traditional approaches in sociology. Finally, these tools have been integrated into a prototype to demonstrate the value of this approach on two case studies and according to different user case scenarios, such as data or annotation browsing or a timeline visualization of opinion groups.

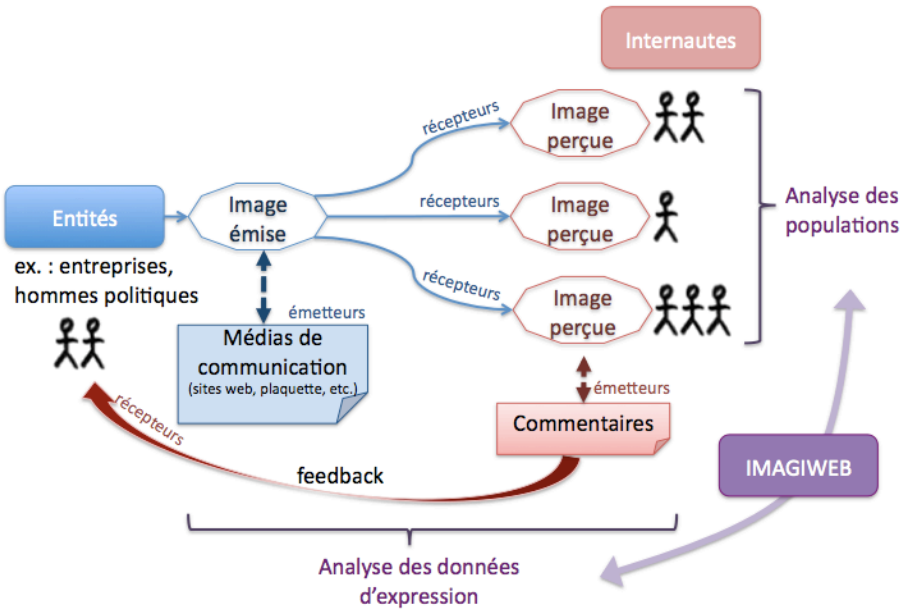
Major Results

This project has shown that it was possible to capture fine-grained opinions about an entity with the help of automatic analysis tools on messages extracted from the Internet. About politicians, we were able to compare with traditional opinion survey results in order to draw conclusions about convergences with our tools or notable differences. In the case of EDF, the implemented prototype has helped prove the conclusion expression by a specialist in semiology, in an exhaustive way, but also to detect new information. From an industrial perspective, this project has helped define a general method to study the reputation on the Internet. This method has been integrated into the AMI Software watch platform.

Scientific Production

The scientific output of this project has been the publication of 6 articles in journals, 21 in conference proceedings and 8 communications. The domain of investigation is primarily in Computer Science (NLDB, DMNLP, TALN), information extraction (ECIR, LREC, CLEF, INEX), text mining (EGC) and social network analysis (SNAM, ASONAM). There was also some communications in national and international conferences (APSA) in political science.
Two programs were implemented during the project (an annotation software and a software for Twitter analysis) and one demonstration prototype was also applied on the two case studies. Two study days were also organized with, in each case, at least 50 attendees.

Illustration



Factual information

The ImagiWeb project was a fundamental research project coordinated by Julien Velcin of the ERIC Lab, University of Lyon 2. In addition, it involved the CEPEL of the University of Montpellier, the LIA of the University of Avignon et des Pays du Vaucluse, the Xerox company, EDF R&D and the company AMI Software. The project began in April 2012 and lasted 42 months. The grant given by the ANR was 840,625 € for a global cost of 3,142,564 €.