

Projet ANR-2012-CORD-002-01

ImagiWeb

Programme CONTINT 2011

Résumé public (français)

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)	1 ^{er} 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)	<ul style="list-style-type: none">- 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 : capturer l'opinion en ligne avec une approche informatique et sociologique

Objectif général du projet et problématique : capturer l'opinion exprimée au sujet d'une entité en combinant des algorithmes développés en informatique et une logique de panélisation sociologique

L'image de nombreuses entités (par ex. célébrités, entreprises, marques) nous parvient principalement par l'intermédiaire de l'existence virtuelle qu'elles mènent sur le Web et dans les nouveaux médias. L'objectif du projet ImagiWeb est d'analyser l'opinion exprimée dans les messages postés sur Internet au sujet de ces entités, à l'aide de techniques informatiques et statistiques, et de la relier aux caractéristiques sociales des individus qui les ont produits en suivant une logique de panélisation novatrice. A cette approche résolument pluridisciplinaire s'ajoute la volonté d'apprécier l'opinion en regard de cibles qui décrivent l'entité (par ex. : les soutiens de l'homme politique ou la politique tarifaire de l'entreprise) et de la suivre de manière dynamique. Ce type de travaux peut avoir un impact aussi bien technique (développer des nouveaux algorithmes et logiciels), stratégique (apporter une nouvelle solution pour la gestion de la réputation en ligne) que sociétal (mieux comprendre la naissance et la diffusion des représentations).

Méthodes et technologies utilisées : combiner une analyse fine des opinions extraites depuis les médias sociaux, un regroupement automatique en catégories et une logique de panélisation

Pour résoudre cette problématique, nous avons choisi de combiner des outils avancés de traitement automatique de la langue (approche linguistique), de fouille de textes (approche statistique) et d'apprentissage automatique (supervisé et non supervisé). La prédiction des cibles et des polarités d'opinion est obtenue à l'aide d'une méthode hybride et active de classification supervisée, tandis que le regroupement des groupes d'opinion homogène est construit à l'aide de clustering probabiliste évolutionnaire. Les producteurs des messages analysés sont identifiés à l'aide d'une stratégie originale de panélisation des internautes mise en place en modernisant une approche suivie traditionnellement en sociologie. Enfin, ces outils sont intégrés dans un prototype permettant de démontrer l'intérêt de l'approche sur deux cas d'étude et selon plusieurs scénarios d'usage envisagés, tels que la navigation dans les données et les annotations ou la visualisation temporelle des groupes d'opinion.

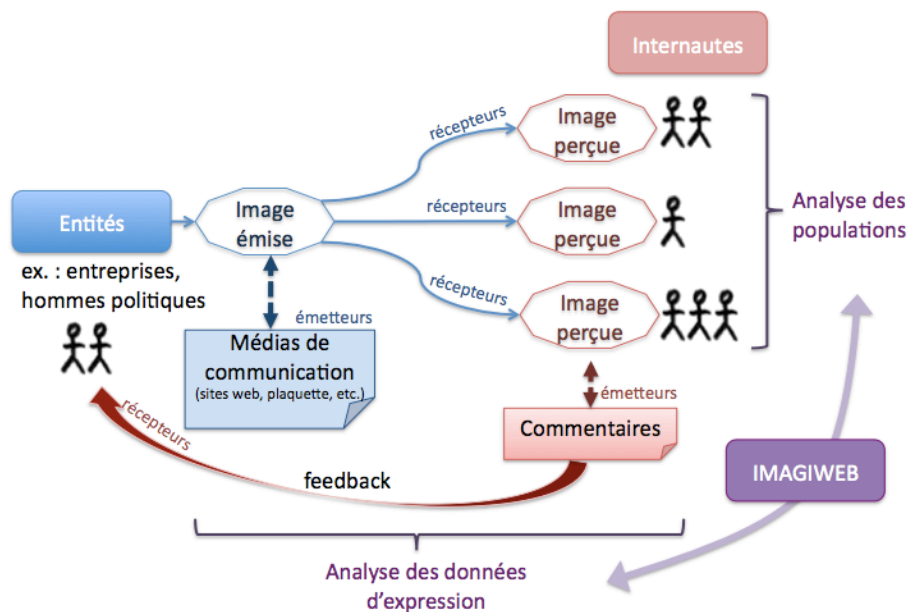
Résultats majeurs du projet

Le projet a principalement permis de montrer qu'il était possible de capturer les opinions fines au sujet d'une entité en utilisant des outils d'analyse automatique des messages d'expression sur Internet. Sur le cas des hommes politiques, il a été possible de comparer cette opinion en ligne aux baromètres habituels de l'opinion et d'en tirer des conclusions sur certaines convergences observées mais surtout sur des différences très marquées. Sur le cas de l'entreprise EDF, le prototype mis en place a permis de confirmer des conclusions tirées par les sémiologues, et ce de manière exhaustive, mais également de proposer des informations nouvelles. D'un point de vue industriel, ce projet a permis l'élaboration d'une méthode générale pour étudier l'image de marque (réputation) sur Internet. Celle-ci a été d'ores et déjà intégrée à la plateforme de veille éditée par AMI Software.

Production scientifique et brevets depuis le début du projet

Le projet a conduit à la publication de 6 articles dans des journaux, 21 dans des actes de conférences et à 8 communications. Les domaines concernés se situent principalement en informatique, notamment en traitement automatique du langage naturel (NLDB, DMNLP, TALN), recherche d'information (ECIR, LREC, CLEF, INEX), fouille de données (EGC) et analyse des réseaux sociaux (SNAM, ASONAM). On remarquera également des communications dans les congrès nationaux et internationaux (APSA) de sciences politiques. Deux logiciels ont été produits durant le projet (un logiciel d'annotation et un logiciel d'analyse de Twitter) et un prototype de démonstration sur les deux cas d'étude traités. Deux journées d'étude ont été organisées, rassemblant chaque fois environ 50 participants.

Illustration



Informations factuelles

Le projet ImagiWeb était un projet de recherche fondamentale coordonné par Julien Velcin du laboratoire ERIC, Université de Lyon 2. Il associait aussi le CEPEL de l'Université de Montpellier, le LIA de l'Université d'Avignon et des Pays du Vaucluse, la société Xerox, EDF R&D et l'entreprise AMI Software. Le projet a commencé en avril 2012 et a duré 42 mois. Il a bénéficié d'une aide ANR de 840 625 € pour un coût global de 3 142 564 €.