

Cette fiche permettra de présenter le projet et son avancement de façon très synthétique sur le site web de l'ANR. Les auteurs autorisent l'ANR à publier le contenu de ce résumé sur son site web ou sur d'autres supports.

Cette communication vise un public scientifique large, il faut donc privilégier une rédaction pédagogique et éviter les explications visant uniquement les spécialistes du domaine.

S'aider de l'exemple fourni.

## Identification du projet

|   |  |
|---|--|
| Acronyme                                    | ImagiWeb   |
| Titre                                       | Images sur le Web : Analyse de la dynamique des images sur le Web 2.0  |
| Programme – Edition                         | ANR CONTINT 2011   |
| Référence ANR                               | ANR-2012-CORD-002-01   |
| Contact coordinateur (Nom, partenaire, mél) | VELCIN Julien, Laboratoire ERIC, <a href="mailto:Julien.Velcin@univ-lyon2.fr">Julien.Velcin@univ-lyon2.fr</a>  |
| Partenaires (société, organismes, labos)    | AMI Software, EDF France, ERIC (Université de Lyon), CEPPEL (Université de Montpellier I), LIA SFR Agorantic (Université d'Avignon et des Pays du Vaucluse), Xerox Research Center Europe (XRCE) |
| Date de début<br>Date de fin                | 1 <sup>er</sup> Avril 2012 – 31 Mars 2015  |
| Pôles de compétitivité                      | Cap Digital, Imaginove, Solutions Communicantes Sécurisées   |
| Coût complet                                | 3 142 564 euros  |
| Aide ANR                                    | 872 197 euros (complément attribué par les pôles compris)  |
| Site web                                    | <a href="http://mediamining.univ-lyon2.fr/people/velcin/imagiweb/">http://mediamining.univ-lyon2.fr/people/velcin/imagiweb/</a>  |
| Date de mise à jour                         | 27 Sept. 2013  |

## Titre d'accroche du projet (1 ligne)

Etude informatique et sociologique des images circulant sur Internet

## Sous-titre / Argument du projet (2 à 4 lignes)

Combiner une approche informatique et sociologique permettra non seulement d'exhiber les images qui circulent sur Internet, mais également de comprendre quelles sont les populations qui se cachent derrière l'expression des opinions constituant ces images.

## Titre de la partie Enjeux & objectifs (1 ligne)

Dévoiler les mécanismes de production et de diffusion des images

## Enjeux & objectifs (20 lignes max)

De nombreuses entités ont une existence virtuelle sur la toile : films, hommes politiques, entreprises, marques, etc. L'enjeu du projet est de dévoiler les mécanismes qui procèdent à la production, la diffusion, l'évolution des opinions des internautes relativement à ces entités, ce que nous appelons les images. Pour cela, l'idée est de mêler une approche informatique, mettant notamment en œuvre des techniques d'analyse automatique des textes d'expressions

postés sur Internet (blogs, tweets), et une approche sociologique afin de déterminer l'identité des producteurs d'opinion. Techniquement, le projet donnera lieu à l'élaboration d'un prototype qui devra permettre de visualiser et d'interroger les images des entités étudiées dans le projet : l'image de N. Sarkozy et de F. Hollande entre les deux tours de l'élection présidentielle de 2012, et l'image de l'entreprise EDF sur le sujet du nucléaire. Concernant les retombées pour la société, un tel outil pourra permettre de mieux comprendre les processus de création et de diffusion des opinions. Ces processus jouent un rôle majeur, entre autres choses, sur l'étude de l'e-reputation des entités qui peuplent le Web.

### **Titre de la partie Méthodes / Approches (1 ligne)**

Combiner une approche informatique et sociologique pour extraire et suivre les images

### **Méthodes / Approches (20 lignes max)**

L'un des principaux atouts du projet est de combiner une analyse automatique des messages textuels produits sur le Web pour en extraire les images, et une étude sociologique pour caractériser au plus près les émetteurs des opinions qui constituent ces images.

L'analyse automatique des textes contenant les opinions est un verrou qui pourra tirer partie d'annotations manuelles reflétant la connaissance des experts. Cette analyse sera réalisée en utilisant à la fois de techniques issues du traitement automatique des langues, permettant une analyse fine des textes (prise en compte de la syntaxe, des négations, des co-références, etc.), mais également de techniques issues des statistiques, afin de proposer des nouvelles méthodes hybrides pour l'extraction des opinions. Les images, résultat de l'agrégation des opinions émises dans les messages, seront reconstruites à l'aide d'algorithmes d'apprentissage automatique issus de l'intelligence artificielle. Ces algorithmes devront prendre en compte le caractère évolutif de ces images. Une fois ces images (re)constituées, différentes techniques seront mises en place pour les interroger de manière conviviale : système de question-réponse, visualisation, navigation en ligne, etc.

La représentativité de ces images est un verrou important qui sera attaqué en proposant de nouvelles méthodes pour caractériser les sources des opinions. Pour cela, il est prévu de construire des panels d'internautes représentatifs afin de suivre dans le temps les opinions d'individus socialement identifiés. Ces panels permettront de collecter les publications émises par les individus et de les analyser au fil du temps en connaissant les caractéristiques sociales de leurs auteurs.

### **Résultats (20 lignes max)**

En ce qui concerne l'Informatique, les retombées principales consistent en l'élaboration de nouveaux modèles et algorithmes pour manipuler les opinions et les images qui circulent sur Internet. Il est prévu d'élaborer un prototype logiciel afin d'interroger ces images de manière conviviale : présentation résumée des images, système de question-réponse, navigation en ligne. Dans un cadre de veille informationnelle, il devra permettre de mieux comprendre l'opinion des internautes exprimée au sujet d'entités de nature diverse (dans nos deux cas d'étude : deux hommes politiques, et une entreprise).

En ce qui concerne les Sciences Humaines et Sociales (SHS), et en particulier la Sociologie, la réussite du projet permettra de redéfinir la manière dont les sociologues définissent et exploitent des panels sur le Web. Il sera ainsi possible de caractériser les populations d'internautes qui expriment les opinions constitutives des images.

Nous soulignons que le projet ImagiWeb a été labellisé par trois pôles de compétitivité : Cap Digital (région Ile-de-France), Imaginove (région Rhône-Alpes) et SCS (région PACA).

### **Perspectives (10 lignes max)**

Les perspectives de ce projet sont nombreuses. D'un point de vue applicatif, il peut donner lieu à l'élaboration de nouveaux outils pour la veille d'information sur Internet. En particulier,

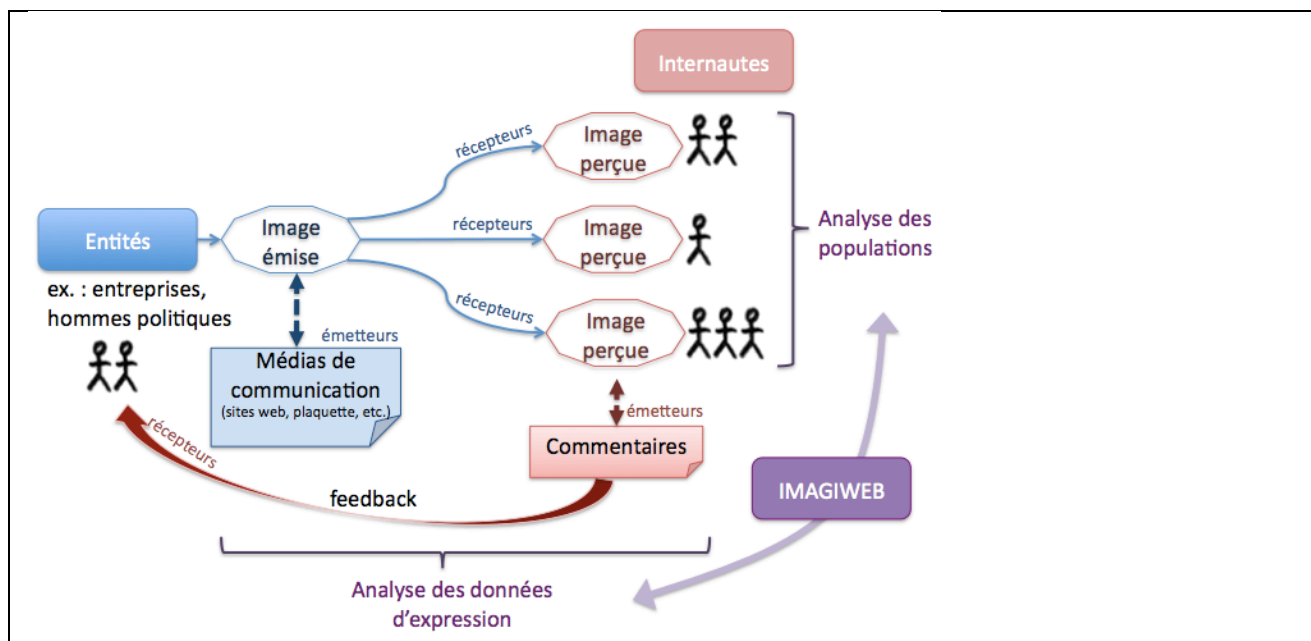
la question de l'image des individus, publics ou privés, des entreprises, des marques... (ce que l'on appelle aujourd'hui l'e-réputation), est une préoccupation majeure dans notre société numérique. Le projet cherche à aller plus loin sur l'analyse des opinions, notamment en découpant l'image des entités en composantes plus fines (que nous appelons des cibles), mais surtout en essayant de retrouver l'identité des producteurs de ces opinions. Pour les sociologues, il s'agit de développer de nouveaux modes d'interrogation des populations qui pourront être utilisés dans bien d'autres contextes d'étude.

### Productions scientifiques et brevets (10 lignes max)

Deux publications scientifiques ont été réalisées en lien avec ImagiWeb (les membres du projet sont soulignés) :

- Structuring typical evolutions using Temporal-Driven Constrained Clustering, M.A. Rizoïu, J. Velcin, S. Lallich (ICTAI 2012). Un nouvel algorithme est proposé pour réaliser un apprentissage automatique adapté à l'analyse d'entités dans la description évolutive dans le temps. Cet algorithme pourra être testé pour reconstruire l'image des entités manipulées dans le projet (hommes politiques, entreprise EDF).
- Extracting Celebrities from Online Discussions, M. Forestier, J. Velcin, A. Stavrianou, D.A. Zighed (ASONAM 2012). Il s'agit d'une étude prospective sur l'extraction de rôles à partir de médias sociaux. Notre intuition est que le rôle d'un internaute peut être déterminant dans la constitution, et surtout dans la diffusion, des images.

### Illustration



Le projet ImagiWeb consiste à mettre en relation une analyse des messages exprimant des opinions et une analyse des populations d'internautes à la source de ces opinions. Les opinions exprimées s'agrègent pour constituer l'une des images possibles de l'entité étudiée, multi-facette par nature, fonction du groupe à la source de l'opinion. Enfin, l'image, qu'elle soit exprimée par l'entité elle-même ou par les internautes, évolue au fur et à mesure du temps. Il faut donc être capable d'en capter au plus près la dynamique.