**ECOS Nord**

***É****valuation-orientation de la* ***CO****opération* ***S****cientifique*

*Programme de coopération et de partenariat*

*universitaire et scientifique*

*France - Amérique Latine*

*(Colombie – Mexique – Venezuela)*

**COOPÉRATION SCIENTIFIQUE AVEC LE MEXIQUE**

**21ème APPEL À PROJETS (clôture : le 16 mai 2014)**

1. Partenaires institutionnels au Mexique : ANUIES (Association nationale des universités et institutions d’éducation supérieure), coordonnateur, CONACYT (Conseil national des sciences et technologies) et SEP (Secrétariat à l’éducation publique).
2. Date limite de réception des projets (formulaire de candidature et document annexes) : le **16 mai 2014**, délai de rigueur. Les projets doivent parvenir en deux exemplaires papier (un original + une copie, voir adresse page suivante) sous couvert du chef d’établissement ainsi que sous format électronique à l’adresse ecos.nord@univ-paris13.fr

Le partenaire mexicain doit déposer son projet simultanément (dans les mêmes termes mais en langue espagnole) auprès du comité mexicain, en utilisant le formulaire fourni par l’ANUIES, qui émet le même appel à propositions.

1. Les projets sont rédigés en français et peuvent porter sur tous les champs de la connaissance. Toutefois, les domaines prioritaires dans le cadre de la coopération franco-mexicaine sont les suivants : changement climatique, protection des écosystèmes et de la biodiversité, catastrophes naturelles, énergies propres et renouvelables et consommation durable, développement technologique (manufacture de haute technologie, nanomatériaux et nano-technologies, matériaux avancés et ingéniéries), santé (prévention des addictions, problèmes de santé publique, médecine préventive, bio-ingéniérie), sécurité alimentaire (production alimentaire, biotechnologies), développement urbain durable, technologies de l’information et de la communication, sciences sociales (lutte contre la pauvreté, économie de la connaissance, économie digitale, migrations, organisations collectives et sécurité urbaine), automobile, aérospatiale.
2. Le Comité scientifique ECOS Nord est réparti en six grands champs disciplinaires. Mettre le chiffre **1** dans la case correspondante au domaine dans lequel votre projet sera évalué. Si vous souhaitez qu’il soit aussi examiné dans un autre domaine, mettre le chiffre **2** dans la case correspondante.

**[ ]  Agronomie, environnement et biologie animale** **[ ]  Sciences physiques et chimiques**

 **(2)[x]  Sciences humaines et sociales**  **[ ]  Sciences de la vie et de la santé**

 **(1)[x]  Mathématiques, informatique et automatique [ ]  Sciences de la terre et de l’univers**.

1. Si plusieurs établissements sont impliqués dans le projet, ils désignent en concertation un seul responsable scientifique. Dans tous les cas, c’est l’établissement de rattachement du responsable de projet qui soumet le dossier au Comité ECOS Nord.
2. Chaque responsable ne peut déposer qu’un seul projet par appel à propositions. S’il a déjà bénéficié d’une aide ECOS Nord pour un projet antérieur, ce dernier doit être terminé et avoir été soumis à une évaluation finale. Le programme ECOS n’a pas vocation à fournir des financements récurrents.
3. Le porteur du projet doit être habilité à diriger des recherches et en activité au moment du dépôt du dossier. Il ne peut y avoir qu’un seul porteur de projet pour la partie française.
4. Procédure de sélection : évaluation en parallèle des projets, en France par des experts du comité ECOS Nord, et au Mexique par leurs homologues mexicains, sélection commune lors d’un comité mixte tenu au dernier trimestre de l’année en cours.
5. Type d’actions retenues : collaboration à un projet scientifique commun intégrant un volet de formation au niveau doctoral.
6. IMPORTANT : Pour être recevable, le projet doit impérativement faire participer un doctorant mexicain. La participation d’au moins un doctorant dans l’équipe française entre dans les critères de sélection. Les bénéficiaires des missions juniors doivent être nommément précisés dans le dossier.
7. Durée des actions retenues :4 ans. Les missions et les stages sont attribués chaque année (du 1er janvier au 31 décembre) et ne peuvent être reportés d’une année sur l’autre.
8. Un même projet peut être présenté en coopération avec deux équipes relevant de deux pays relevant du programme ECOS. Dans ce cas, les porteurs français du projet répondent à l’appel à propositions des deux pays concernés et indiquent dans chaque dossier qu’il s’agit d’un projet tripartite.
9. Nature des soutiens attribués par ECOS Nord : missions pour des chercheurs confirmés en activité et inscrits dans le projet initial ; stages pour des étudiants en cours de thèse. ECOS Nord assure le coût du transport jusqu’au lieu de la mission au Mexique pour les Français et celui du séjour en France pour les Mexicains, sur une base de réciprocité.
10. Les moyens attribués sont notifiés au début de chacune des quatre années du projet. Les demandes de moyens doivent être justifiées de façon précise. À titre indicatif, les projets retenus ont bénéficié jusqu’à présent de deux missions, une junior et une senior, dans les deux sens, chaque année.
11. Les moyens apportés par le présent programme sont destinés à favoriser les mouvements de chercheurs. Le financement des recherches proprement dites ainsi que celui des bourses doctorales relèvent d’autres institutions.
12. Il appartient aux chercheurs français, notamment aux responsables de projets, de prendre toutes les dispositions utiles afin d’assurer **la protection efficace et la répartition correcte de la propriété intellectuelle**.
13. Tutelles du Comité ECOS Nord : ministère des affaires étrangères, ministère de l’enseignement supérieur et de la recherche.
14. Nous vous invitons à consulter la page <http://www.anuies.mx/media/docs/convocatorias/pdf/ecos_2014.pdf> sur le site Internet de l’ANUIES et le site des comités ECOS [www.univ-paris13.fr/cofecub-ecos/](http://www.univ-paris13.fr/cofecub-ecos/)

**POUR RÉPONDRE À CET APPEL À PROJETS :**

**1 - TÉLÉCHARGER LA FICHE-PROJET DE CANDIDATURE SUR LE SITE D’ECOS :**

**http://www.univ-paris13.fr/cofecub-ecos/**

**2 - L’ADRESSER, UNE FOIS COMPLÉTÉE, EN FRANCAIS ET EN 2 EXEMPLAIRES\*, DONT UN ORIGINAL SOUS COUVERT DU CHEF CHEF D’ÉTABLISSEMENT, À L’ADRESSE SUIVANTE:**

**COMITÉ ECOS Nord / Bureau A113A**

**UNIVERSITÉ PARIS 13 / 99 avenue Jean-Baptiste Clément**

**93430 VILLETANEUSE**

(Tél : 01 49 40 44 40 Fax : 01 49 40 20 27)

**3 -ADRESSER ÉGALEMENT UNE VERSION ELECTRONIQUE À l’ADRESSE E-MAIL DU COMITÉ :**

**ecos.nord@univ-paris13.fr**

\*Y joindre obligatoirement en **deux exemplaires** les ***curriculum vitae*** des responsables français et mexicains et ceux, plus succints, des autres chercheurs et doctorants français et mexicains inscrits dans le projet.

 Vous pouvez également joindre éventuellement un exemplaire du dernier rapport d’activité de votre unité.

**Note sur une clause particulière du programme ECOS Nord franco-mexicain**

**Bourses doctorales pour les universitaires mexicains**

 Le renforcement des corps académiques des universités mexicaines est une priorité pour le Mexique. À cet effet, l'accord SEP-CONACYT-ANUIES-ECOS Nord stipule que chaque projet présenté doit obligatoirement prévoir la formation doctorale d’un étudiant mexicain. Cette formation doit s'effectuer au sein des systèmes nationaux d'enseignement supérieur français et mexicain dans le cadre du projet.

Deux programmes mexicains de bourses facilitent le formation doctorale des enseignants concernés au niveau de la maestria et du doctorat :

* **programme PROMEP** (*programa de mejoramiento del profesorado*) de la *Subsecretaria de Educatión Superior e Investigación Científica* (SESIC) de la *Secretaria de Educación Pública* (SEP)*.*

Plus d’informations sur le site http://promep.sep.gob.mx

* *programa de becas crédito al extranjero* du **CONACyT :** les universitaires faisant partie des centres SEP-CONACYT ou d'autres institutions d’enseignement supérieur peuvent solliciter une bourse auprès du Conacyt. Plus d’informations sur le site www.conacyt.mx

**Bénéficiaires :**

Universitaires mexicains d'une institution d'enseignement supérieur. Les études doctorales peuvent se réaliser à temps partiel en alternance en France et au Mexique ou à temps plein en France, soit en codirection (le doctorat est celui de l'établissement français ou celui de l’établissement mexicain, selon l'inscription du doctorant), soit en cotutelle (le doctorat est celui de l'établissement français et celui de l’établissement mexicain suite à une convention entre les établissements selon l'arrêté du 18 Janvier 1994).

Pour être éligible, le candidat à une bourse mexicaine doit remplir les conditions requises par le programme auprès duquel il postule.

Les candidats doivent cependant être attentifs aux dates d’échéance des appels d’offres concernant ces programmes de bourse.

**COOPÉRATION SCIENTIFIQUE AVEC LE MEXIQUE**

**21ème APPEL À PROJETS : FORMULAIRE DE CANDIDATURE**

(à adresser à ECOS Nord avec les documents annexes, au plus tard le **16 mai 2014**, le cachet de la poste faisant foi).

**Important : trois pages d'informations spécifiques précèdent cette fiche-projet**.

***- Remplissez les cadres sans modifier la taille ni la disposition des zones de texte.***

***- Merci de développer les sigles non usuels que vous utilisez.***

***- Pour une meilleure lisibilité des sections 1 à 7, écrivez en caractères gras vos réponses.***

Nom du responsable du projet français **:** **Aldo Gangemi**

Nom du responsable du projet mexicain **: Héctor Jiménez Salazar**

 **1. Intitulé du projet**

Titre (en gras) : **Cartographie de la diaspora des savoirs par fouille sémantique des résultats de recherches sur le web**

Mots-clés (4 maximum) / Discipline **: recherche d’expert, migrations hautement qualifiés, web sémantique, *machine reading*, *web mining*, traitement automatique des langues**

 **2. Établissements**

 **3. Responsables du** **projet**

 **NB : Le responsable français doit être habilité à diriger des recherches et en activité à la date de dépôt du projet.**

**En France :**

Nom et prénom **: GANGEMI Aldo** Grade **: Professeur**

Adresse administrative **: Institut Galilée - Université Paris-Nord, 99, avenue Jean-Baptiste Clément, 93430 Villetaneuse**

|  |  |  |
| --- | --- | --- |
|  Téléphone **: +33 1 49 40 40 81** | Télécopie **: +33 1 48 26 07 12** | Mél: aldo.gangemi@lipn.univ-paris13.fr |

**Au Mexique :**

Nom et prénom **: Héctor Jiménez Salazar** Grade **: Professeur**

Adresse administrative **: Universidad Autónoma Metropolitana, Unidad Cuajimalpa, Av. Vasco de Quiroga 4471, Col. Santa Fé, Cuajimalpa, Delegación Cuajimalpa de Morelos, México, D.F., C.P.05300**

|  |  |  |
| --- | --- | --- |
|  Téléphone **: +52 5591 77 6650 x 6950** | Télécopie **: +52 5591 77 6650 x 6950** | Mél **:** hgimenezs@gmail.com |

|  |  |  |
| --- | --- | --- |
|  Téléphone **:**  | Télécopie **:**  | Courriel **:**  |

**En France :**

Établissement **: Université Paris 13**

Laboratoire ou équipe (auquel appartient le responsable du projet)**: Laboratoire d’Informatique de Paris Nord**

Statut et numéro de l’unité  : UMR 7030 CNRS

*(ex : UMR 999 CNRS)*

Nom du Directeur : Laure Petrucci

**Au Mexique :**

Établissement **: Universidad Autónoma Metropolitana (UAM), unidad Cuajimalpa**

Laboratoire (ou équipe) **: Departamento de Tecnologías de la Información**

|  |  |  |
| --- | --- | --- |
|  Téléphone **:**  | Télécopie **:**  | Courriel **:**  |

 **4. Liste des chercheurs participant au projet** :

\* Commencer par le porteur du projet et indiquer obligatoirement les doctorants qui y participent, même s’ils sont mentionnés par ailleurs. Chaque équipe doit comporter au minimum 3 chercheurs ou ingénieurs de recherche. Seuls les chercheurs figurant sur cette liste peuvent bénéficier ultérieurement de missions.

\* Rappel : les chercheurs participant à un projet ECOS Nord doivent être en activité au moment du dépôt du projet.

|  |  |  |  |
| --- | --- | --- | --- |
| **Nom** | **Diplôme universitaire****le plus élevé \*\*** | **Établissement actuel** **d'affectation** | **Fonction actuelle et/ou****diplôme poursuivi** |
| **En France :** |  |  |  |
| **Aldo GANGEMI****Davide BUSCALDI** | **Doctorat, Roma-Sapienza,****1994****Doctorat, Politécnica de****Valencia, 2010** | **LIPN/Université Paris 13****LIPN/Université Paris 13**  | **Professeur****Maître de Conférences** |
| **Jorge GARCIA FLORES**  | **Doctorat, Paris-Sorbonne,** **2007** | **LIPN/Université Paris 13** | **Ingénieur de recherche** |
| **Ehab Hassan** | **Master 2, Paris Dauphine,****2011** | **LIPN/Université Paris 13** | **Doctorant** |
| **Dana Diminescu**  | **Doctorat** | **ENST – Télécom ParisTech** | **Maître de Conférences** |
|  |  |  |  |
|  |  |  |  |
| **Au Mexique :** |  |  |  |
| **Héctor Jiménez Salazar** | **Doctorat,**  | **UAM-Cuajimalpa** | **Enseignant chercheur** |
| **Esaú Villatoro Tello** | **Doctorat** | **UAM-Cuajimalpa** | **Enseignant chercheur**  |
| **David Eduardo Pinto** **Avendaño** | **Doctorat** | **BUAP** | **Enseignant chercheur**  |
|  |  |  |  |
| **Raymundo Jacobo****Jiménez Muñoz** | **Master** | **UAM-Cuajimalpa** | **Étudiant**  |
| **Darnes Vilariño Ayala** | **Doctorat, BUAP** | **BUAP** | **Enseignant chercheur** |

 **5. Résumé du projet (ne pas dépasser la limite inférieure du cadre, le projet sera détaillé section 8)** :

 **6. Soutiens publics et privés obtenus (PCP, CNRS, Union européenne, industriels, etc.)**

 **se rattachant au projet**

**La localisation des migrants hautement qualifiés est un problème crucial pour les pays émergents. Au delà des méthodes propres à la sociologie de la migration, comme les enquêtes de terrain ou les recensements de la population, une solution attractive est la mise en place des techniques informatiques de fouille sémantique de textes avec le but de cartographier la trajectoire professionnelle des membres de la diaspora des savoirs à partir de documents disponibles sur le web. Cette tâche présente plusieurs défis pour les technologies du langage naturel: la recherche et extraction d'information biographique; la désambigüisation des noms propres, termes scientifiques et toponymes; le calcul de similarité sémantique entre des fragments courts de texte produits par un moteur de recherche; la classification terminologique par domaine de connaissance ainsi que l'extraction d’évènements. Notre objectif est double: d'un côté il s'agit de produire des cartographies individuelles pour la trajectoire professionnelle (géographique et thématique) des membres de la diaspora des savoirs à partir des sources de données institutionnelles (comme par exemple la base d'ex-boursiers CONACYT); d'un autre côté nous envisageons de permettre la recherche d'experts appartenant à la diaspora hautement qualifiée à partir de mots clés (par exampe: chercher un expert en freinage à courants de Foucault d'origine mexicaine ayant été formé en France). L'expérience conjointe des équipes de recherche composant le consortium offre un large spectre méthodologique pour approfondir la collaboration entre partenaires: Dans un premier temps nous mettrons en place un cadre expérimental avec des données issues de sources institutionnelles (CONACYT, CONICET, COLCIENCIAS) et de projets de recherche en sociologie de la diaspora des savoirs (CIDESAL). Dans un deuxième temps nous étudierons de manière critique les méthodes linguistiques et statistiques les plus aptes à la fouille sémantique des résultats d'un moteur de recherche. Cette étape permettra le classement des résultats par *clustering* et sa désambigüisation par WSI (*Word Sense Induction*) pour en extraire les relations sémantique entre un nom propre, des institutions, des lieux géographiques et des dates constituant une trajectoire professionnelle. Une attention particulière sera donnée à la caractérisation épistémologique du domaine d'expertise en faisant appel à des ontologies des savoirs et à des calculs de proximité sémantique. Cet ensemble impliquera la représentation de relations extraites sous la forme d'un graphes RDF construit par *machine reading*, d'où la cartographie sera générée. Notre but ultime est d'établir les bases méthodologiques pour une nouvelle génération d'outils de recherche d'experts qui prennent en compte le contexte biographique et géographique propre à chaque personne.**

\* En France et au Mexique, de manière succincte (exemple : PCP, 2007-2010) ; les titres complets et détails seront donnés en section 8.6.

Ancien projet ECOS ? OUI – NON

Si oui, indiquez le code, l’année et le nom du projet :

Autres soutiens publics dans le passé **:**

En cours **:**

Soutiens privés :

 **7. Formation doctorale**

**Indiquez, sachant que le projet doit impérativement faire participer au moins un doctorant mexicain, les étudiants mexicains et français envisagés dans le cadre du projet, en mentionnant obligatoirement les bénéficiaires, l'organisme finançant la bourse doctorale et la nature du doctorat. Une attention particulière sera portée aux doctorats en cotutelle au sens de l'arrêté du 18 Janvier 1994 (convention de cotutelle inter-établissements).**

*a - Etudiants mexicains et organisme auquel la demande de bourse a été faite* **4** *:*

 **Raymundo Jacobo JIMÉNEZ MUÑOZ a fini ses études de Master 2 dans la Faculté de Sciences Informatiques de la BUAP (Mexique) sur la conception et l’implémentation d’un système de recherche sémantique d’information spécialisé en traitement de documents sur des gisements pétrolifères. Le profil de l’étudiant est parfaitement adapté aux besoins de ce projet, puisque il a été formé en Traitement Automatique des Langues et Ontologies. Il est en train de demander une bourse du *Programa de becas crédito CONACYT* (gouvernement mexicain) dans l’appel 2014.**

*b - Etudiants français et organisme auquel la demande de bourse a été faite* 4 *:*

 **Ehab Hasssan est déjà boursier doctoral.**

 EhaEhab Hassanm

 **8. Description détaillée du projet**

**Dans cette dernière partie, on respectera la succession des sections, mais on pourra librement ajuster leur taille, intercaler entre elles autant de pages que nécessaire et utiliser des annexes. Il n’est pas utile de frapper le texte en caractère gras.**

**N.B. : les projets doivent être rédigés en termes strictement équivalents en français et en espagnol, des rédactions différentes pouvant être sources d'appréciations différentes de la part des experts des deux comités.**

**8.1 Identification du sujet et état des connaissances**

La localisation des migrants hautement qualifiés est un problème crucial pour les pays émergents. Au-delà des sources des données démographiques, comme les recensements de population, les enquêtes de terrain sur la population active, les sources gouvernementales (base des permis de travail, bases de visas) et les registres de la [OECD, 2002] , une solution attractive est la mise en place des techniques informatiques de fouille sémantique de textes avec le but de cartographier la trajectoire professionnelle des membres de la diaspora des savoirs à partir exclusivement d’information publiée sur le web.

La première application des technologies du langage à la sociologie de la migration a été le système *Unoporuno[[1]](#footnote-1)* [Garcia-Flores et al. 2012], développé dans le cadre du projet CIDESAL ([[2]](#footnote-2)). Ce système permet de classer un des membres potentiels de la diaspora selon son mouvement migratoire : le mouvement linéaire (la personne est partie à l’étranger pour se former et n’est jamais revenue) ; le mouvement circulaire (la personne est revenue à son pays d’origine) et l’absence de mouvement (la personne n’est jamais partie à l’étranger). Il reçoit en entrée une liste des noms en contexte (par exemple, des fiches bibliographiques) et fouille le web pour classer les personnes par apprentissage automatique. Grâce aux membres du consortium CIDESAL, une base de13000 membres potentielle de la diaspora latino-américaine des savoirs a pu être constituée. Ses membres proviennent des base des publications scientifiques (Web of Science) ainsi que des bases d’anciens boursiers des organismes scientifiques de l’Argentine (CONICET), la Colombie (COLCIENCIAS), l’Uruguay (SNI) et le Mexique (CONACYT). [CIDESAL, 2014]*.* La principale limite de cette approche est qu’elle ne fait aucune extraction des données, mais se limite au classement automatique par traits sémantiques présents dans les résultats des recherches automatiques sur le web.

Quant à cartographier la diaspora à travers les donnés disponibles sur le web, l’approche phare est le *e-Diaspora Atlas*, où 8000 sites web produits où fréquentés par les émigrants du monde ont été analysés par 80 chercheurs pour caractériser les interactions propres à chaque diaspora nationale à partir de l’étude des topologies crées par les hyperliens *[[3]](#footnote-3)* [Diminescu 2008].

Une autre approche de cartographie de la diaspora, le système Name-sor[[4]](#footnote-4) se base sur une étude onomastique de l’origine des noms propres des utilisateurs de *Twitter* pour cartographier la densité des différentes diasporas nationales sur *Twitter* [Shokhenmayer 2010].

Les systèmes de recherche d’experts sur le web (*expert finding*) n’ont jamais été directement appliqués à la recherche des émigrés hautement qualifiés sur le web. En fait, les données biographiques sont peu ou pas de tout considérés par les méthodes de recherche d’expert, qui se concentrent sur la recherche des liens épistémiques entre un expert et son domaine d’expertise. Pourtant les passerelles entre la recherche d’experts et la recherche de personnes sur le web (*Web People* Search) nous paraissent abondantes et fertiles. La recherche d’experts gagnerait en précision si elle prenait en compte le contexte biographique des experts (pays d’origine, pays de formation, parcours professionnel) et réciproquement le domaine de spécialité (ou toute simplement la profession) serait un trait distinctif pour la désambigüisation d’homonymes dans la recherche de personnes. Deux approches sortent du lot dans cette famille d’applications : *Saffron* [Monaghan et al. 2010] et *Sig.ma* [Tummarello et al. 2010]. Par rapport aux méthodes, Saffron s’appuie sur du TAL pour l’annotation des pages web pertinentes, tandis que *Sig.ma* exploite les métadonnées sémantiques. Toutes les deux font abondant usage des représentations sémantiques en RDF pour caractériser épistémologiquement le domaine de spécialité des experts.

Les meilleurs résultats rapportés dans la littérature en fouille sémantiques de recherches sur le web ont été obtenus avec la méthode d’induction des sens des mots (*Word Sense Induction)* [Di Marco & Navigli 2013] qui repose sur des couches de cooccurrence lexicale et l’usage de *DBpedia* et *Babelnet* pour la désambigüisation des entités recherchées. Cette méthode n’a jamais été appliquée à la fouille d’experts sur le web, mais d’autres méthodes de fouille sémantiques l’ont été, sans que ces outils soient encore disponibles. Quant à la découverte des relations sémantiques propres à une trajectoire professionnelle, plusieurs solutions existent dans la littérature, come *Text Runner* [Yates et al. 2007], un extracteur d’information ouverte capable de parcourir un texte, annoter les parties du discours, les syntagme nominaux et en en extraire un ensemble de couples censés représenter des relations sémantiques. De même, *FRED* [Gangemi et al. 2014] est un « lecteur automatique » (*machine reader*) qui est capable d’extraire d’un texte la majeur quantité possible d’information (entités nommées, sens, taxinomies, relations, événements) et de transformer le tout en un graph RDF où les données extraites ont un lien à une sources des données publique comme *DBpedia*.

**8.1.1 Originalité**

L’originalité de notre approche repose sur le fait d’appliquer des méthodes de pointe en TAL, fouille sémantique du web et *machine reading* au problème bien ancrée dans la réalité : celui de la recherche d’experts appartenant à la diaspora des savoirs. L’apport des résultats escomptés aurait un impacte épistémologique double : en technologies sémantiques il permettrait de produire une nouvelle génération d’outils en recherche d’experts qui prennent en compte le contexte biographique des experts ; en sociologie ils permettraient de produire une nouvelle source des données possible pour l’étude des migrations hautement qualifiés. Au delà de l’objectif particulier de cartographier de façon individuelle la trajectoire géographique et thématique des hautement qualifiés, il serait intéressant de chercher une corrélation statistique entre les experts de notre corpus et les tendances nationales rapportées par les organismes internationaux comme l’OCDE et l’ONU. Ce résultat permettrait de répondre à la question posée par les méthodes numériques : jusqu’à quelle point la réalité sur le web permet de répondre à des questions sur la réalité du terrain ? [Rogers 2010]

Un mot est ici nécessaire sur l’originalité de notre consortium. Le LIPN possède une expertise approfondie en méthodes combinées de traitement automatique des langues et de représentation de connaissances. Cette vocation à produire des représentations sémantiques à partir des annotations textuelles s’est vue renforcé avec l’intégration du système *FRED* et des méthodes de *machine reading* aux activités quotidiennes de recherche du labo. Cette expérience en représentation de phrases courtes en langage naturelle sous forme des graphes RDF/OWL, tout en faisant appel à des ressources structurés comme *DBPedia* comme base de connaissances extralinguistiques, permet d’envisager un traitement sémantique poussé pour les résultats de recherches sur le web considérés comme pertinents dans la trajectoire professionnelle d’une personne. De plus, le labo possède également une bonne connaissance en production massive de requêtes et fouille des résultats des moteurs de recherche, puisque la toute dernière version du système *Unoporuno* a été produite au sein du LIPN. La participation annuelle de l’équipe à la tâche de similarité sémantique dans la campagne SEMEVAL permet aussi d’envisage l’intégration aisée des techniques de calcul de proximité sémantique aux méthodes de *clustering* des résultats. Quant à la pertinence sociologique des cartographies des diasporas des savoirs produites, l’équipe de l’ENST Télécom ParisTech, pionnière dans ce domaine, est incontournable. De son côté, les équipes mexicaine de la UAM-Cuajimalpa et la BUAP (Puebla) possèdent une expérience approfondie en classification par *clustering* et désambigüisation de SMS qui peut très bien s’adapter aux textes courts caractéristiques des moteurs de recherche. L’équipe de la UAM-Cuajima a beaucoup travaillé sur l’extraction d’information géographique, étape toute aussi incontournable dans la production des cartographique à partir de données biographiques publiées sur le web.

**8.2 Objectifs**

* Développer des méthodes computationnelles qui permettent de cartographier automatiquement la trajectoire professionnelle des membres la diaspora des savoirs de manière individuelle. De même, cette méthode permettra de retrouver des experts à partir des compétences spécifiques tout en prenant compte du contexte biographique de l’expert (example : trouver un expert en freinage par courants de Foucault d’origine mexicaine et formé en France)
* Établir une séquence de traitement automatique des langues et *machin*e *reading* qui permette de passer d’un texte court à une représentation structurée en RDF en faisant appel à des méthodes d’apprentissage automatique et de *clustering* par WSI [Di Marco & Navigli 2013]. Cette représentation sera la base pour cartographier la trajectoire des experts.
* Agréger les cartographies individuelles pour en dégager des tendances à niveau national et calculer la corrélation statistiques entre les statistiques issues du web et celles publiées par les organismes internationaux comme l’OCDE ou l’ONU.

**8.2.1 Objectif scientifiques**

* Prendre les corpus et les ressources méthodologiques développés par des projets en sociologie des diasporas hautement qualifiées (en particulier CIDESAL) comme base empirique pour la construction des cartographies individuelles des experts.
* Fouilles sémantiquement les contextes sémantiquement riches des noms des personnes (par exemple : un fiche bibliographique où une patente) pour en extraire des organisations, des toponymes et des sujets de recherche ou d’expertise qui puissent être combinés au nom pour générer des requêtes web à caractère biographique pour un moteur de recherche.
* Intégrer l’outil de *deep machine reading* FRED [Etzioni et al. 2006 ; Gangemi et al. 2014] au système de recherche des membres de la diaspora hautement qualifiée *Unoporuno* [Garcia Flores et al. 2012] pour étendre ses fonctionnalités au cas d’utilisation classique de la recherche d’experts (*expert finding*) [Bordea & Buitelaar 2012].
* Renforcer la découverte des éléments sémantiques dans les résultats des moteurs de recherche par des méthodes d’apprentissage automatique.
* Évaluer les cartographies individuelles des experts de la diaspora hautement qualifiée avec le corpus et les méthodes proposés par CIDESAL.
* Agréger les trajectoires individuelles dans des statistiques qui permettent, par exemple, de déterminer quelles ont été les principaux pays d’accueil pour la diaspora hautement qualifié du Mexique où le taux d’émigration par domaine d’expertise.
* Calculer la corrélation statistique entre nos données issues du web et les chiffre du terrain de l’OCDE et l’ONU.

**8.2.2 Objectifs institutionnels**

* Renforcer la coopération bilatérale entre des groupes scientifiques mexicains et français.
* Promouvoir la coopération à plus long terme entre institutions de deux pays.
* Entamer une synergie entre les groupes de recherche interdisciplinaire où les expert en TAL et les chercheurs des études sociales travaillent ensemble.
* Former des chercheurs et des doctorants dans les méthodes numériques appliquées à la sociologie.
* Renforcer les programmes d’enseignement au niveau Master et Doctorat au Mexique avec des thèmes de TAL développés tout au long du projet.

**8.3 Méthodologie**

**8.2.2 Méthodologie scientifique**

Dans un premier temps on structurera les théories, les concepts et les techniques et les idées autour de l’idée d’une cartographie individuelle d’un expert pour en dégager un état de l’art exhaustif et une évaluation des principales ligne d’action avec le but d’établir un projet faisable dans les délais prévus.

Dans un deuxième temps il sera nécessaire de concevoir des cadres empiriques pour les hypothèses proposées, notamment en ce qui concerne l’annotation des résultats des moteurs de recherche et l’extraction des relations biographiques permettant de reconstruire une trajectoire professionnelle. On évaluera avec les mesures classiques en traitement automatique des langues : rappel, précision, F-mesure, précision moyenne, exactitude, taux d’erreur, ainsi que des mesures de corrélation entre nos résultats et ceux rapportés dans la littérature.

Dans un troisième temps on établira une critique des méthodes et techniques développés pour confirmer ou infirmer nos hypothèses à la lumière des résultats empiriques par rapport à la construction des cartographies des experts appartenant à la diaspora hautement qualifiée.

En fin, on évaluera les avantages et les limites des méthodes et techniques développés au cours du projet, ainsi que des extensions possibles à d’autres types de problématiques.

Donc, voici nos trois hypothèses :

1. **Il est possible d’établir une cartographie de la trajectoire épistémique d’un expert ainsi que de sa mobilité professionnelle et géographique à partir de l’analyse des données (non structurées, quasi-structurées et structurées) publiées sur le web.**
2. **Le modèle de cartographie de la diaspora hautement qualifiée ici proposé est applicable au problème de la recherche d’experts, c’est-à-dire, à la construction d’une liste de noms d’experts à partir de certaines aptitudes et compétences recherchées.**
3. **Il existe une corrélation entre les statistiques de mobilité des hautement qualifiés calculées exclusivement à partir des données extraites du web et celles rapportées par les organismes internationaux tels que l’OCDE et l’ONU.**

**8.2.2 Méthodologie technologique**

1. Construction d’un socle empirique de tests (*testbed*) composé d’un échantillon statistiquement significatif des experts appartenant à la diaspora hautement qualifiée
2. Actualisation de l’outil de recherche des diasporas hautement qualifiées *Unoporuno* pour l’utiliser comme instrument d’observation des diasporas sur le web.
3. La sortie actuelle d’*Unoporuno* sera considérée comme une référence minimale de comparaison (*baseline*) pour mesurer l’efficacité de la nouvelle approche.
4. Expériences et évaluation des diverses possibilités méthodologiques et technologiques de chaines de TAL, fouille sémantique et apprentissage automatique pour déceler la chaine de traitement optimale.
5. Modélisation du champ épistémologique d’expertise en se basant sur les techniques courantes en recherche d’experts [Aleman-Meza et al. 2007 ; Li et al. 2006]. Expériences et évaluation.
6. Intégration d’*Unoporuno* dans la chaine de transformation d’annotations sémantique en représentations RDF avec FRED. Cette étape représente un défi majeur en traitement multilingue étant donné que FRED traite seulement des textes en langue anglaise et qu’*Unoporuno* travaille avec des données multilingues propres aux parcours multinationaux des hautement qualifiés. Il faudra envisager alors une étape translinguistique où les analyses en anglais soient « transposées » à l’espagnol, l’allemand et le français [Banea et al. 2008].
7. Évaluation du processus d’extraction des informations minimales pour reconstruire une trajectoire professionnelle : origine, formation élémentaire, formation spécialisée, domaine d’expertise, événements marquants.
8. Structuration des données sous forme de graphes RDF
9. Visualisation cartographique des trajectoires professionnelles avec Cytoscape[[5]](#footnote-5) où Cartodb[[6]](#footnote-6)
10. Agréger les données statistiques pour construire des cartographies à plus grande échelle.
11. Calculer la corrélation entre les statistiques issues du web et celles issues du terrain (OCDE, ONU)

**8.3.4 Références**

[Aleman-Meza et al. 2007] Boanerges Aleman-Meza , Uldis Bojars , Harold Boley , John G. Breslin , Malgorzata Mochol , Lyndon Jb Nixon , Axel Polleres , Anna V. Zhdanova. Combining RDF vocabularies for expert finding. In Proceedings of the 4th European Semantic Web Conference (ESWC2007), number 4519 in Lecture Notes in Computer Science, (2007).

[Banea et al. 2008] Carmen Banea, Rada Mihalcea, Janyce Wiebe and Samer Hassan, Multilingual Subjectivity Analysis Using Machine Translation, in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008), Honolulu, Hawaii, October 2008.

[CIDESAL, 2014] CIDESAL : Construction d’Incubateurs des Diasporas du Savoir en Amérique Latine.

Rapport final d’activité. EuroAid/Commission Européenne, 2014 *(in press)*

[Bordea & Buitelaar 2012] Georgeta Bordea, Paul Buitelaar. Expertise Mining. Proceedings of the 21st National Conference on Artificial Intelligence and Cognitive Science, Galway, Ireland, 2010.

[Di Marco & Navigli 2013] Antonio Di Marco, Roberto Navigli: Clustering and Diversifying Web Search Results with Graph-Based Word Sense Induction. Computational Linguistics 39(3): 709-754 (2013)

[Diminescu 2008] Diminescu D. The connected migrant: an epistemological manifesto. Social Science Information (December 2008) vol. 47 no. 4 565-579

[Etzioni et al. 2006] Oren Etzioni, Michele Banko and Michael J Cafarella. Machine Reading. AAAI Conference on Artificial Intelligence, 2006.

[Gangemi et al. 2014] Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero. Frame-based detection of opinion holders and topics: a model and a tool. IEEE Computational Intelligence, 9(1), 2014

[Garcia Flores et al. 2012] J. J. Garcia Flores, Pierre Zwigenbaum, Zhao Yue and W.A. Turner (2012) Tracking Researcher Mobility on the Web Using Snippet Semantic Analysis. In H. Ishara & K. Kanzaki, Advances in Natural Language Processing 8th International Conference on NLP, October 22-24, Kanazawa, Japan. (JapTAL 2012) Springer Lecture Notes in Computer Sciences, vol. 7614, ISBN 978-3-642-33983-7, pp. 180-191.

[Li et al. 2006] Jie Li, Harold Boley, Virendrakumar C. Bhavsar, and Jing Mei. Expert Finding for eCollaboration Using FOAF with RuleML Rules. Montreal Conference on eTechnologies MCTECH, (May 17, 2006)

[Monaghan et al. 2010] Fergal Monaghan, Georgeta Bordea, Krystian Samp, Paul Buitelaar (2010)

Exploring Your Research: Sprinkling some Saffron on Semantic Web Dog Food, 9th International Semantic Web Conference (ISWC), Semantic Web Challenge, Shanghai, China.

[OECD, 2002] International Mobility of the Highly Skilled. Policy Brief, OECD, 2002, Paris.

[Rogers 2010] Richard Rogers, "Internet Research: The Question of Method," Journal of Information Technology and Politics 7 (2010): 241-260.

[Shokhenmayer 2010] Evgeny Shokhenmayer Cartography in onomastic research / TRANS vol.18; Virtual INST World Conference. CCKS: Cities, Cultures, Knowledge Societies. Section ‘Cartographic Science as a Reflection of Knowledge Societies and Cultures’, 2010

[Tummarello et al. 2010] Giovanni Tummarello, Richard Cyganiak, Michele Catasta, Szymon Danielczyk, Renaud Delbru, Stefan Decker "Sig.ma: Live views on the Web of Data", Journal of Web Semantics: Science, Services and Agents on the World Wide Web - Volume 8, Issue 4, November 2010, Pages 355-364

[Yates et al. 2007] Alexander Yates, Michael Cafarella, Michele Banko, Oren Etzioni, Matthew Broadhead, and Stephen Soderland. 2007. TextRunner: open information extraction on the web. In Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations (NAACL-Demonstrations '07)

**8.4 Calendrier du projet**

|  |  |  |
| --- | --- | --- |
| Année | Activités | But |
| 1 | * État de l’art
* *Testbed* à partir des corpus CIDESAL (Conacyt + Conicet + Colciencias)
* Actualsation d’Unoporuno
* Prise en main de FRED
* Charte éthique pour la manipulation de données biographiques extraites du web
* Cas d’utilisation
 | Développer un cadre théorique, méthodologique et technologique d’analyse et d’évaluation de la qualité des données pour le projet.  |
| 2 | * *Clustering* des résultats *d’Unoporuno* par *WSI* (*Word Sense Induction*)
* Ontologie des savoirs / Vocabulaires / compilation de ressources épistémiques appropriées (lexiques, dictionnaires, thésaurus)
* Annotation des entités nommées + évaluation
* Extraction d’évènements + évaluation
* Grammaire formelle pour reconnaitre les noms des personnes en espagnol, français, anglais et allemand + évaluation
 | Chaîne de traitement TAL + Machine reading |
| 3 | * Intégration *Unoporuno+FRED*
* *Crosslingual FRED*
* Production de RDF + évaluation
* Premières cartographies individuelles
 | Développer une chaine multilangue, le graph RDF et les premiéres cartographies individuelles.  |
| 4 | * Graphes agrégées
* Statistiques agrégées
* Correlation entre les données du web et les données du terrain (OCDE)
* Évaluation
 | Cartographies à plus grande échelle. Statistiques a plus grande échelle. Corrélation web vs terrain |

**Actions**

|  |  |
| --- | --- |
| Année | Actions |
| 1 | 1. Une mission de 15 jours pour 1 chercheur mexicain en France.
2. Une mission de 7 jours pour 2 chercheurs mexicains en France
3. Une mission de 15 jours pour 1 chercheur français au Mexique
4. Une mission de 7 jours pour 2 chercheurs français au Mexique
5. Une mission d’un mois pour un doctorant français au Mexique
6. Publication conjointe d’articles de prise de position
 |
| 2 | 1. Une mission de 7 jours pour 3 chercheurs mexicains en France
2. Une mission de 7 jours pour 3 chercheurs français au Mexique
3. Une mission de 7 jours pour 1 doctorant français au Mexique
4. Publication conjointe d’articles avec les premiers résultats
 |
| 3 | 1. Une mission de 7 jours pour 2 chercheurs mexicains en France
2. Une mission de 7 jours pour 2 chercheurs français au Mexique
3. Participation conjointe dans des jurys de mémoires de master et thèses de doctorat
4. Publication conjointe des résultats scientifiques
 |
| 4 | 1. Organisation d’un congrès de présentation des résultats et clôture
2. Participation conjointe dans le jury de thèse du doctorant mexicain financé par CONACYT.
3. Publication conjointe des résultats scientifiques
 |

**8.5 Résultats escomptés au terme de l'action**

**N.B. : toute publication, sur tout support (papier, affiches, film, vidéo…) y compris les résumés (abstracts) produits lors de congrès, devra impérativement mentionner le soutien du programme ECOS/ANUIES pour pouvoir être prise en compte lors de l'évaluation finale du projet.**

* Corpus et batterie de tests pour l’extraction des trajectoires professionnelles du web.
* Méthode pour cartographier les diasporas des savoirs à partir des données disponibles sur le web.
* Outil de production des cartographies individuelles des experts, soit à partir d’un nom d’expert, soit à partir d’un domaine d’expertise et des contraintes biographiques.
* Ontologie des savoirs
* Publications conjointes des équipes du Mexique et de la France.
* Organisation d’un colloque franco-mexicain avec des doctorants et chercheurs intéressés dans la thématique du projet.
* Formation de ressources humaines (étudiants et chercheurs) dans la fouille sémantique du web.

**8.6 Antécédents de coopération**

**Les coopérations éventuellement mentionnées en *section 6* devront être détaillées ici. Joindre en annexe, le cas échéant, les références des publications cosignées, thèses, équipements réalisés, projets conjoints, brevets, etc.**

**8.7 Publications**

**Liste des cinq publications les plus significatives des 2 équipes concernées dans les cinq dernières années (en
relation ou non avec le projet) :**

***Équipe française:***

* **Aldo Gangemi**
* Ciaramita M, Gangemi A, Ratsch E, Rojas I, Saric J. Unsupervised Learning of Semantic Relations between Concepts of a Molecular Biology Ontology, in Proceedings of International Joint Conference on Artificial Intelligence (IJCAI2005), 2005.
* Baldassarre C., Daga E., Gangemi A., Gliozzo A., Salvati A., Troiani G. Semantic Scout: Making Sense of Organizational Knowledge. P. Cimiano and S: Pinto (eds.): Proceedings of the Conference on Knowledge Engineering and Knowledge Management (EKAW2010), Springer, 2010.
* Valentina Presutti, Francesco Draicchio and Aldo Gangemi. Knowledge extraction based on Discourse Representation Theory and Linguistic Frames. A. ten Teije and J. Völker (eds.): Proceedings of the Conference on Knowledge Engineering and Knowledge Management (EKAW2012), LNCS, Springer, 2012.
* Aldo Gangemi. A Comparison of Knowledge Extraction Tools for the Semantic Web. Proceedings of ESWC2013, LNCS, Springer, 2013.
* Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Alberto Musetti and Paolo Ciancarini. Aemoo: exploring knowledge on the Web. Proceedings of WebSci2103 (ACM Web Science Conference), ACM, 2013.
* **Dana Diminescu**
* Diminescu D. (ed.): Special issue: Diasporas on the Web (2012) Social Science Information Journal, December 2012; 51 (4) Publisher: SAGE, on behalf of Maison des Sciences de l’Homme, Paris, ISSN 0539-0184, 245 pages.
* Diminescu D. The connected migrant: an epistemological manifesto. Social Science Information (December 2008) vol. 47 no. 4 565-579
* Diminescu D., 2007 « Le passage par l’écran : ou l’émergence de nouvelles frontières », in Actes du colloque Les frontières de l’Europe, Ed Universitatii, Bucarest, pp 263-274
* Diminescu D., (2006), « L’usage du téléphone portable par les migrants en situations précaire », in Quatre ans de recherche urbaine (2001-2004), ACI-Ville Min de la recherche, E. Bajolet, MF Mattei, JM Rennes (sld ), Presse Universitaire François Rabelais, MSH Ville et territoire, Tours, 2006, pp90-98
* Diminescu D., (2006), « Genèse d’une figure de migrant, in Cosmopolitiques n°11/2006, ed Apogée, Paris, pp. 63-73
* **Davide Buscaldi**
* Davide Buscaldi and Paulo Rosso. A conceptual density-based approach for the disambiguation of toponyms. Int. J. Geogr. Inf. Sci., 22(3) :301–313, 2008
* Davide Buscaldi, Paolo Rosso, José Manuel Gómez, and Emilio Sanchis. Answering questions with an n-gram based passage retrieval engine. Journal of Intelligent Information Systems (JIIS), 34(2) :113–134, 2009.
* Paolo Rosso, Santiago Correa, and Davide Buscaldi. Passage retrieval in legal texts. Journal of Logic and Algebraic Programming, 80(3-5) :139 – 153, 2011.
* Guillaume Cleuziou, Davide Buscaldi, Vincent Levorato, and Gaël Dias. A pretopological framework for the automatic construction of lexical-semantic structures from texts. In CIKM ’11 : Proceedings of the 20th ACM international conference on Information and knowledge management, pages 2453–2456, New York, NY, USA, 2011. ACM.
* Davide Buscaldi. Approaches to disambiguating toponyms. SIGSPATIAL Special, 3(2) :16–19, 2011.
* **Jorge J. Garcia Flores**
* J. J. Garcia Flores, Pierre Zwigenbaum, Zhao Yue and W.A. Turner (2012) Tracking Researcher Mobility on the Web Using Snippet Semantic Analysis. In H. Ishara & K. Kanzaki, Advances in Natural Language Processing 8th International Conference on NLP, October 22-24, Kanazawa, Japan. (JapTAL 2012) Springer Lecture Notes in Computer Sciences, vol. 7614, ISBN 978-3-642-33983-7, pp. 180-191.
* J. Garcia-Flores and W.A. Turner (2011) A Nominal Filter for Web Search Snippets: Using the Web to Identify Members of Latin America’s Highly Qualiﬁed Diaspora. In I.Z. Batyrshin & G. Sidorov, Mexican International Conference on Artificial Inteligence (MICAI 2011), IEEE press, ISBN: 978-1-4577-2173-1, pp. 45-50.
* B. Djioua, J.J. Garcia Flores, A. Blais, J-P. Desclés, G. Guibert, A Jackiewicz., F. Le Priol, L. Nait-Baha, B. Sauzay (2006) EXCOM: an automatic annotation engine for semantic information. In G. Sutcliff & R. Goebel (eds.): Florida Artificial Intelligence Research Society Conference (FLAIRS 2006), Association for the Advancement of Artificial Intelligence Press, ISBN 978-1-57735-261-7, pp. 285-290 .
* J.J. Garcia-Flores. (2004) Semantic Filtering of Textual Requirements Descriptions. In F. Méziane & E. Métais (editores): Natural Language Processing and Informations Systems (NLDB 2004), Springer, Lecture Notes in Computer Science, ISBN 978-3-540-22564-5, pp. 427-434 .

***Équipe mexicaine:***

* **Hector Jiménez Salaza**
* Esaú Villatoro-Tello, Christian Sánchez-Sánchez, Héctor Jiménez-Salazar, Wulfrano Arturo Luna-Ramírez, Carlos Rodríguez-Lucatero: UAM at INEX 2012 Relevance Feedback Track: Using a Probabilistic Method for Ranking Refinement. CLEF (Online Working Notes/Labs/Workshop) 2012
* David Pinto, Darnes Vilariño Ayala, Yuridiana Alemán, Helena Gómez-Adorno, Nahun Loya, Héctor Jiménez-Salazar: The Soundex Phonetic Algorithm Revisited for SMS Text Representation. TSD 2012: 47-55
* David Pinto, Paolo Rosso, Héctor Jiménez-Salazar: A Self-enriching Methodology for Clustering Narrow Domain Short Texts. Computer Journal. 54(7): 1148-1165 (2011)
* David Pinto, Paolo Rosso, Héctor Jiménez-Salazar: On the Assessment of Text Corpora. NLDB 2009: 281-290
* Alberto Bañuelos-Moro, Héctor Jiménez-Salazar, José de Jesús Lavalle-Martínez: Test of complementarity on sentence extraction methods.Procesamiento del Lenguaje Natural 40 (2008)

**• David Eduardo Pinto Avendaño**

* David Pinto, Helena Gómez-Adorno, Darnes Vilariño Ayala, Vivek Kumar Singh: A graph-based multi-level linguistic representation for document understanding. Pattern Recognition Letters 41: 93-102 (2014)
* Paolo Rosso, Marcelo Errecalde, David Pinto: Analysis of short texts on the Web: introduction to special issue. Language Resources and Evaluation 47(1): 123-126 (2013)
* David Pinto, Paolo Rosso, Héctor Jiménez-Salazar: A Self-enriching Methodology for Clustering Narrow Domain Short Texts. Computer Journal. 54(7): 1148-1165 (2011)
* Claudia Denicia-Carral, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda and David Pinto-Avendaño, Bilingual Document Clustering: Evaluating Cognates as Features, Journal of Information and Library Science, 35(3): 265-286, 2011
* David Pinto, Jorge Civera, Alberto Barrón-Cedeño, Alfons Juan, Paolo Rosso: A statistical approach to crosslingual natural language tasks. Journal of Algorithms 64(1): 51-60 (2009)

**• Esaú Villatoro Tello**

* Fernando Sánchez-Vega, Esaú Villatoro-Tello, Manuel Montes-y-Gómez, Luis Villaseñor Pineda, Paolo Rosso: Determining and characterizing the reused text for plagiarism detection. Expert Syst. Appl. 40(5): 1804-1813 (2013)
* Esaú Villatoro-Tello, Antonio Juárez, Manuel Montes-y-Gómez, Luis Villaseñor Pineda, Luis Enrique Sucar: Document ranking refinement using a Markov random field model. Natural Language Engineering 18(2): 155-185 (2012)
* Esaú Villatoro-Tello, Christian Sánchez-Sánchez, Héctor Jiménez-Salazar, Wulfrano Arturo Luna-Ramírez, Carlos Rodríguez-Lucatero: UAM at INEX 2012 Relevance Feedback Track: Using a Probabilistic Method for Ranking Refinement. CLEF (Online Working Notes/Labs/Workshop) 2012
* Esaú Villatoro-Tello, Ricardo Omar Chávez García, Manuel Montes-y-Gómez, Luis Villaseñor Pineda, Luis Enrique Sucar: A Probabilistic Method for Ranking Refinement in Geographic Information Retrieval. Procesamiento del Lenguaje Natural 44 (2010)
* Maya Carrillo, Esaú Villatoro-Tello, Aurelio López-López, Chris Eliasmith, Luis Villaseñor Pineda, Manuel Montes-y-Gómez: Concept Based Representations for Ranking in Geographic Information Retrieval. IceTAL 2010: 85-96

**• Darnes Vilariño Ayala**

* Darnes Vilariño Ayala, David Pinto, Beatriz Beltrán, Saul León, Esteban Castillo, Mireya Tovar: A Machine-Translation Method for Normalization of SMS. MCPR 2012: 293-302
* David Pinto, Helena Gómez-Adorno, Darnes Vilariño Ayala, Vivek Kumar Singh: A graph-based multi-level linguistic representation for document understanding. Pattern Recognition Letters 41: 93-102 (2014)
* David Pinto, Darnes Vilariño, Carlos Balderas, Mireya Tovar and Beatriz Beltran, Evaluating n-gram Models for a Bilingual Word Sense Disambiguation Task, Computación y Sistemas 15(2), 2011
* Darnes Vilariño Ayala, David Pinto, Mireya Tovar, Carlos Balderas, Beatriz Beltrán: A Probabilistic Model Based on n-Grams for Bilingual Word Sense Disambiguation. MICAI (1) 2010: 82-91
* David Pinto, Darnes Vilariño, Carlos Balderas, Mireya Tovar, and Beatriz Beltrán: A Naive Bayes Approach to Cross-Lingual Word Sense Disambiguation and Lexical Substitution. MCPR 2010. LNCS 6256, Springer-Verlag : 352-361

**8.8 Informations complémentaires**

 **Cochez les cases en cas de réponse positive**

**[ ]** Ce projet a été présenté en totalité ou en partie à un autre organisme (indiquer les moyens éventuellement
obtenus) :

**[ ]** Ce projet a été retenu par **:**  pour la période du **:** au **:**

**[ ]** Ce projet n'a été soumis à aucun organisme, même partiellement.

 **[ ]** Le projet a aussi été déposé pour l’appel à projets ECOS en cours avec **la Colombie** / **le Venezuela** **/ le Chili / l’Argentine / l’Uruguay** (rayez ou effacez la mention inutile).

 En cas de réponse positive, le projet doit être déposé également dans le cadre de l’appel à projets du pays concerné.

**8.9 Suggestion de 2 référés**

 **Proposer les noms et adresses postales et électroniques de 2 référents en poste en France qui ne collaborent pas au projet. L'un d'eux sera sollicité, en plus d'un expert choisi par le Comité.**

**Pierre Zweigenbaum**

**pz@limsi.fr**

**(+33) 01 69 85 80 04**

**LIMSI-CNRS
BP 133, F-91403 ORSAY CEDEX**

**Jean-Baptiste Meyer**

[**jean-baptiste.meyer@ird.fr**](jean-baptiste.meyer%40ird.fr)

**Tél : +33(0)4 13 55 07 47**

**Université de Provence, case 10**

**3, place Victor Hugo**

**13331 Marseille, cedex 03 France**

**AVIS ET VISA DU**

**CHEF D'ÉTABLISSEMENT N° 1**

(voir page suivante pour établissement n° 2)

 **Cette feuille devra être remise en même temps que le projet avant la date de clôture de l'appel à projets.**

Date **:**

Nom, prénom et signature du **responsable français** du **projet :**

Avis, nom, prénom et signature du **responsable** de **l'unité de recherche** à laquelle appartient le responsable du projet :

Adresse administrative :

Avis, nom, prénom, cachet et signature du **chef d'établissement5** ou de son représentant dûment habilité **:**

Adresse administrative :

Cachet et signature précédée de la mention "par délégation" s’il y a lieu :

**5**Pour une université, il s'agit de son Président. Les UFR ne sont pas des établissements. Pour tout autre organisme de recherche et établissement d'enseignement supérieur, il s'agit du Directeur Général ou de son représentant dûment habilité.

**AVIS ET VISA DU**

**CHEF D'ÉTABLISSEMENT N° 2**

 **Les équipes mixtes appartenant à deux organismes (ou plus), en particulier le CNRS, l'INSERM, l’IRD et l'INRA, pourront, si elles souhaitent leur patronage, faire remplir l’avis ci-dessous par un second chef d'établissement. Il devra être renvoyé au secrétariat du comité ECOS Nord dans un délai de deux semaines après la clôture de l'appel à projets. Il sera inutile d'y joindre une nouvelle copie du projet.**

Date **:**

Nom, prénom et signature du **responsable français** du **projet :**

Avis, nom, prénom et signature du **responsable** de **l'unité de recherche** à laquelle appartient le responsable du projet :

Adresse administrative :

Avis, nom, prénom, cachet et signature du **chef d'établissement 5** ou de son représentant dûment habilité **:**

Adresse administrative :

Cachet et signature précédée de la mention "par délégation" s’il y a lieu :

**5**Pour une université, il s'agit de son Président. Les UFR ne sont pas des établissements. Pour tout autre organisme de recherche et établissement d'enseignement supérieur, il s'agit du Directeur Général ou de son représentant dûment habilité.

1. http://unoporuno.limsi.fr/unoporuno [↑](#footnote-ref-1)
2. Construction d’Incubateurs des Diasporas du Savoir en Amérique Latine : <http://bit.ly/1nLmhzo> [↑](#footnote-ref-2)
3. [http://www.e-diasporas.fr](http://www.e-diasporas.fr/)/ [↑](#footnote-ref-3)
4. <http://namesorts.com> [↑](#footnote-ref-4)
5. <http://www.cytoscape.org/> [↑](#footnote-ref-5)
6. <http://cartodb.com> [↑](#footnote-ref-6)