

Conception et mise en œuvre d'un outil d'évaluation de la qualité de services de reconnaissance automatique d'entités nommées basé sur l'API WIND

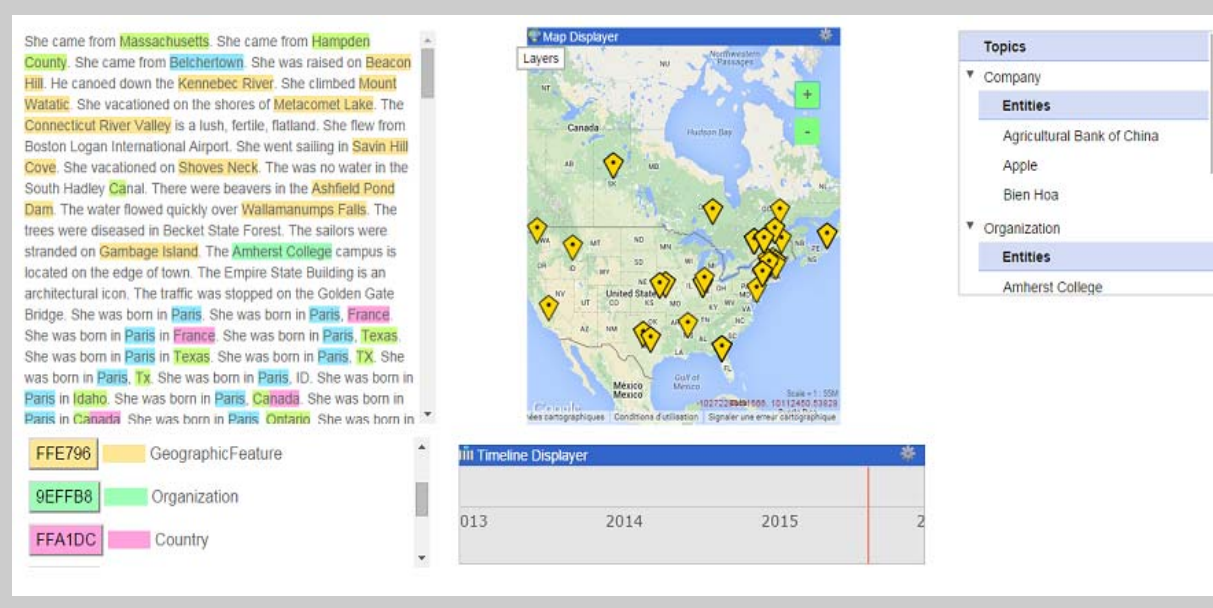
Contact : Marquesuzaà Christophe, Christophe.Marquesuzaa@iutbayonne.univ-pau.fr

Description du sujet

L'évaluation de la qualité des services de reconnaissance d'entités nommées (REN) est un problème important qui se pose à la communauté scientifique en Recherche d'Information (RI). Ces services permettent de repérer automatiquement dans des textes (en anglais et en français) des lieux, des dates ou des thèmes cités dans ces textes ; ces résultats sont restitués sous la forme de fichiers XML, JSON...

Sur la base de prototypes existants, les enseignants-chercheurs de l'équipe T2I souhaitent évaluer différents services de REN tels que Alchemy API, TextRazor, GéoPot ou encore DBpedia Spotlight.

Voici ci-dessous une copie-écran du prototype existant permettant de visualiser les résultats d'annotation lors de l'appel d'un service de REN :



The screenshot displays the T2I prototype interface, which is used for visualizing Named Entity Recognition (NER) results. The interface is divided into several sections:

- Text Annotation:** A paragraph of text is shown with various entities highlighted in different colors. A legend on the left identifies the colors: FFE796 for GeographicFeature, 9EFFB8 for Organization, and FFA1DC for Country.
- Map Display:** A map of the United States and surrounding regions (Canada, Mexico) is shown. Yellow diamond markers indicate the locations of the annotated entities. The map includes a scale bar and a coordinate system.
- Timeline Display:** A horizontal timeline is shown at the bottom, with years 2013, 2014, 2015, and 2016. A red vertical line indicates the current date.
- Topics Panel:** A panel on the right side of the interface lists topics under the heading "Topics". It includes sections for "Company" (with "Entities" listed: Agricultural Bank of China, Apple, Bien Hoa) and "Organization" (with "Entities" listed: Amherst College).

L'objet du stage est d'améliorer et étendre cet outil afin non seulement de visualiser des résultats retournés par plusieurs services de REN (affichage des lieux sur une carte, présentation des dates sur une frise chronologique, etc.) mais aussi de faciliter la fusion des résultats de l'invocation de différents services de REN.

D'un point de vue conceptuel, il s'agira de critiquer/étendre le modèle sous-jacent de représentation des données géographiques extraites par les outils de REN. Ce travail pourra aussi faire l'objet une proposition de publication de recherche.

D'un point de vue opérationnel, il s'agira d'améliorer ou de mettre au point plusieurs modules pour notamment :

- *Invoyer* facilement des services de REN et de récupérer leurs résultats ;
- *Visualiser* les résultats retournés par des ces services de REN sur une interface graphique ;
- *Evaluer* la pertinence des résultats retournés par un service de REN donné ;
- *Comparer* les résultats retournés par plusieurs services de REN ;
- *Fusionner* ces résultats.

Les interactions possibles entre ces différents modules devront être prises en compte au moment de définir l'architecture du prototype final.

Prérequis / Technologies

Ce sujet s'adresse aux étudiants en Master 2 Informatique, ou élève de dernière année d'une École d'ingénieur en informatique. Vous êtes rigoureux dans votre travail mais aussi créatif avec une forte envie d'apprendre.

Les technologies web telles que JavaScript, PHP, les services web, XML et JSON doivent être maîtrisées. Deux frameworks tels que Bootstrap (pour l'IHM) ou Symfony (pour l'organisation modulaire) devraient être pris en main.

Nous fournissons l'API JavaScript WIND (*Web Interaction Design*) pour la visualisation des données spatiales et temporelles. Elle est documentée, dispose de tutoriels (<http://erozate.iutbayonne.univ-pau.fr/windapi/>) et repose sur JavaScript, Ajax, API Open Layers, API Google Maps, API SIMILE Timeline ou encore XML.

Lieu du stage et gratification

Le stage, d'une durée du stage est entre 4 et 6 mois, se déroulera au LIUPPA (Laboratoire d'Informatique de Pau et des Pays de l'Adour) dans l'équipe T2I (Traitement des Informations et des Interactions). Géographiquement, le stage s'effectuera sur le site de la côte basque à Anglet. Ce stage sera gratifié sur toute sa durée.

Candidature

Le dossier de candidature doit contenir :

- le CV
- les relevés de notes des 2 dernières années (M1 et M2)
- la lettre de motivation
- tout autre document pouvant appuyer la candidature

Références

- *Conception d'applications web géographiques guidée par les contenus et les usages : cadre méthodologique et opérationnalisation avec l'environnement WINDMash*. Patrick Etcheverry, Sébastien Laborie, Christophe Marquesuzaà, Thierry Nodenot, The Nhan Luong. Journal d'Interaction Personne-Système ([JIPS](#)), Volume 3, 42 pages. 2014
- *Modélisation centrée utilisateur final appliquée à la conception d'applications interactives en géographie : une démarche basée sur les contenus et les usages*. Thèse de doctorat de The Nhan Luong (<http://tel.archives-ouvertes.fr/tel-00775888/>) soutenue le 12 décembre 2012 à l'Université de Pau et des Pays de l'Adour.
- *Évaluation de la pertinence des résultats en recherche d'information géographique. Définition d'un cadre expérimental et validation de l'apport des dimensions de l'information géographique*. [Guillaume Cabanac](#), [Damien Palacio](#), Christian Sallaberry, [Gilles Hubert](#). Revue [Document Numérique](#) 14(2): 169-191 (2011)
- *Une expérience de fusion pour l'annotation d'entités nommées*. Caroline Brun, Nicolas Dessaigne, Maud Ehrmann, Baptiste Gaillard, Sylvie Guillemin-Lanne, et al. (<https://hal-upec-upem.archives-ouvertes.fr/hal-00621594>) 16^e Conférence sur le Traitement Automatique des Langues Naturelles (TALN'09), Juin 2009, France. 10 pp., 2009.
- Prototypes existants