HAL
open science

Un modèle et une plateforme pour l'évaluation et la comparaison de services de reconnaissances d'entités nommées spatiales

Christophe Marquesuzaà, Patrick Etcheverry, Christian Sallaberry

To cite this version:

Christophe Marquesuzaà, Patrick Etcheverry, Christian Sallaberry. Un modèle et une plateforme pour l'évaluation et la comparaison de services de reconnaissances d'entités nommées spatiales. SAGEO'2016, Dec 2016, Nice, France. hal-01620619

HAL Id: hal-01620619

https://hal.science/hal-01620619v1

Submitted on 20 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Un modèle et une plateforme pour l'évaluation et la comparaison de services de reconnaissances d'entités nommées spatiales

Christophe Marquesuzaà , Patrick Etcheverry , Christian Sallaberry

UNIV PAU & PAYS ADOUR, LABORATOIRE D'INFORMATIQUE DE L'UNIVERSITE DE PAU ET DES PAYS DE L'ADOUR, EA3000, 64600, ANGLET, France

christophe.marquesuzaa@iutbayonne.univ-pau.fr,patrick.etcheverry@iutbayonne.univ-pau.fr,christian.sallaberry@univ-pau.fr

RÉSUMÉ. Cet article s'intéresse à l'évaluation des systèmes de reconnaissance d'entités nommées spatiales. Nous présentons un état de l'art des principaux systèmes connus et nous mettons en évidence la difficulté et le coût pour évaluer et comparer la qualité des résultats produits par ces systèmes. Pour répondre à ce problème nous nous appuyons sur le cadre d'évaluation SNERBM et nous proposons un modèle dédié à l'évaluation des entités nommées. Nous présentons ensuite une plateforme intégrant ce modèle et permettant de mettre en oeuvre un processus d'évaluation conforme au cadre SNERBM. Nous concluons cet article en présentant les premiers résultats d'expérimentation et nous ouvrons quelques perspectives pour faire évoluer le processus d'évaluation et la plateforme support.

ABSTRACT. This paper focuses on the evaluation of spatial named entities recognition systems (SNERS). We present an overview of the main knonw systems and we highlight the difficulty and the cost to evaluate and to compare the quality of the results produced by these systems. To overcome this problem, we use an evaluation framework named SNERBM and we propose a model dedicated to the evaluation of the named entities recognized by these different SNERS. Next, we present a platform integrating this model and carrying out an evaluation process according to the SNERBM framework. We conclude this paper by presenting the first experimentation results and we open perspectives for both the evaluation process and the underlying platform.

MOTS-CLÉS : Entités nommées spatiales, cadre d'évaluation, modèle de représentation d'entités nommées, plateforme d'évaluation.

KEYWORDS: Spatial named entities, evaluation framework, named entities representation model, evaluation platform

1. Introduction

La reconnaissance d'entités nommées (REN) dans des textes (Grishman, Sundheim, 1996 ; Chinchor, 1998), consiste à identifier des syntagmes appelés entités nommées (noms propres, expressions de temps et expressions numériques, par exemple) catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, dates, quantités, distances, valeurs, acronymes, abréviations, etc. Parmi les entités nommées, les noms de lieux, que nous appellerons entités nommées spatiales (ENS), désignent des objets géographiques tels que des entités administratives (commune, par exemple), des éléments du relief, des éléments hydrographiques, etc.

Des systèmes de reconnaissance d'entités nommées spatiales (SRENS) sont très largement utilisés, mais souvent sans en connaître les forces et faiblesses. C'est pourquoi nous utilisons le cadre d'évaluation SNERBM (*Spatial Name Entity Recognition BenchMark*) comme référentiel commun (Palacio *et al.*, 2016). Ce cadre permet l'évaluation de l'efficacité (*effectiveness*) de SRENS. Il permet également d'envisager le choix d'un système, ou encore la combinaison de différents systèmes, particulièrement adaptés aux catégories d'entités nommées spatiales (ville, barrage, montagne, par exemple) majoritairement présentes dans un corpus donné.

Dans cet article, nous proposons un modèle dédié à l'évaluation d'entités nommées retournées par des SRENS. Nous proposons également une plateforme s'appuyant sur ce modèle et permettant l'évaluation de SRENS de manière semi-automatique. L'objectif à terme est de pouvoir invoquer différents SRENS pour analyser un même texte et obtenir automatiquement une évaluation de la qualité des résultats retournés.

Pour faciliter cette évaluation, nous proposons un processus semi-automatique qui permet, à partir d'un fichier résultat de référence construit par des experts, de mesurer automatiquement les taux de rappel, de précision et de F1-Mesure (Manning *et al.*, 2008) des entités reconnues par divers SRENS. Au delà d'une évaluation chiffrée, le modèle et la plateforme proposée doivent également permettre aux experts de visualiser graphiquement les résultats obtenus afin de faciliter leur interprétation.

Cet article est organisé comme suit. La section 2 recense différents SREN, le type d'entités qu'ils sont capables d'annoter, les campagnes d'évaluation associées et le coût de mise en place de ces campagnes. Aprés avoir rappellé les principes de l'évaluation SNERBM en section 3, nous proposons en section 4 un modèle étendu d'entités nommées dédié à la visualisation d'annotations et à l'évaluation automatique de SREN spatiales et de SREN temporelles. Dans la section 5, nous proposons ensuite une plateforme intégrant le modèle précédent et permettant d'évaluer de manière semi-automatique des SRENS. La section 6 présente les premières expérimentations menées sur la plateforme. Enfin, nous concluons par un bilan et des perspectives sur nos prochains travaux.

2. Etat de l'art

Un grand nombre de SREN ont été proposés dans la littérature lors des vingt dernières années (Lieberman *et al.*, 2010 ; Lingad *et al.*, 2013). Le Tableau 1 propose une liste non exhaustive de ces systèmes.

Tableau 1. Liste non exhaustive de SREN et de campagnes d'évaluation

Système de REN	Annotation			URL	Campagnes d'évaluation
	Spat.	Temp.	Thém.		
Abner			✓	pages.cs.wisc.edu/~bsettles/abner/	(Marrero *et al.*, 2009)
Alchemy	✓		✓	www.alchemyapi.com/	(Rizzo *et al.*, 2012)
Annie	✓	✓	✓	gate.ac.uk/sale/tao/splitch6.html#x9-1300006	(Marrero *et al.*, 2009)
CiceroLite	✓	✓	✓	www.languagecomputer.com/products/text-annotation/cicerolite.htm	
Clavin	✓			clavin.bericotechnologies.com/	(D'Ignazio, 2013)
Dbpedia Spotlight	✓		✓	github.com/dbpedia-spotlight/dbpedia-spotlight/wiki/Web-service	(Rizzo *et al.*, 2012)
FreeLing	✓		✓	nlp.lsi.upc.edu/freeling/node/1	(Marrero *et al.*, 2009)
FOX	✓		✓	139.18.2.164:4444/demo/index.html#!/home	
Geolocator	✓			github.com/geoparser/geolocator-3.0	(Gelernter, Zhang, 2013)
Gisgraphy	✓			www.gisgraphy.com/documentation/user-guide.htm#geocodingwebservice	
Heideltime		✓		code.google.com/p/heideltime/	
LingPipe	✓		✓	alias-i.com/lingpipe/	(Marrero *et al.*, 2009)
Metacarta	✓			www.metacarta.com/	
Mondeca	✓	✓	✓	www.mondeca.com/	
NERD	✓	✓	✓	nerd.eurecom.fr/documentation	
OpenCalais	✓		✓	www.opencalais.com	(D'Ignazio, 2013)
Poolparty	✓		✓	www.poolparty.biz/	
Saplo	✓		✓	saplo.com/	(Rizzo *et al.*, 2012)
Supersense	✓	✓	✓	sourceforge.net/projects/supersensetag/	(Marrero *et al.*, 2009)
TextPro	✓		✓	textpro.fbk.eu/	(Marrero *et al.*, 2009)
TextRazor	✓		✓	www.textrazor.com/named_entity_recognition	
Time Bliography	✓	✓		spaceandtime.wsiabato.info/tGIS.html	
Unlock	✓		✓	edina.ac.uk/unlock/texts/getstarted.html	(Tobin *et al.*, 2010)
Yago Naga Aida	✓		✓	www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/webservice/	
Yahoo! Content Analysis	✓		✓	developer.yahoo.com/search/content/V2/contentAnalysis.html	(Rizzo *et al.*, 2012)
Yahoo! PlaceSpotter	✓			developer.yahoo.com/boss/geo/docs/key-concepts.html	(D'Ignazio, 2013)
YooName	✓		✓	yooname.wordpress.com/	(Marrero *et al.*, 2009)
YQL	✓			developer.yahoo.com/yql/guide/	
Zemanta	✓	✓	✓	www.zemanta.com/	(Rizzo *et al.*, 2012)

Aussi, de nombreuses campagnes d'évaluation proposent des plateformes pour évaluer des systèmes de REN ou de recherche d'information géographique (RIG) (Leidner, 2007 ; Andogah, 2010 ; Nouvel, 2012) :

- MUC : Message Understanding Conference (Chinchor, 1998),
- MET : Multilingual Extraction Task (Chinchor, 1998),

– IREX : Information Retrieval and Extraction eXercise (Sekine, Eriguchi, 2000),

– CoNLL : Computational Natural Language Learning conference (Tjong Kim Sang, De Meulder, 2003),

– HAREM : Avaliaçao de sistemas de Reconhecimento de Entidades Mencionadas (Santos *et al.*, 2006),

– GeoClef : Geographic Cross Language Evaluation Forum (Mandl *et al.*, 2009),

– ACE : Automatic Content Extraction program (Strassel *et al.*, 2008),

– EVALITA : Evaluation of NLP and Speech Tools for Italian (Lenzi *et al.*, 2013),

– TREC-CS : Text Retrieval Conference – Contextual Suggestion (Voorhees, 2001 ; Dean-Hall *et al.*, 2013).

Cependant, les cadres d'évaluation mis en œuvre reposent sur des corpus et des métriques différents et, en définitive, peu de SREN du Tableau 1 y ont participé. Par conséquent, il est impossible à ce jour de connaître les performances relatives de SREN suivant un même référentiel.

D'autres approches, dédiées à l'évaluation d'un SREN particulier ou d'un groupe de SREN de même domaine, utilisent des corpus plus petits dont les ressources annotées ne sont que rarement mises à disposition. Bucher *et al.* (2005) ont ainsi proposé d'adapter des techniques d'évaluation existantes pour évaluer le système de RIG SPIRIT. Marrero *et al.* (2009) présentent une plateforme qui a permis l'évaluation des systèmes de REN Supersense, Abner, Annie, FreeLing, TextPro, YooName, ClearForest et Lingpipe. Tobin *et al.* (2010) décrivent une approche dédiée à l'évaluation de système de RENS et plus particulièrement des modules de désambiguïsation et de géolocalisation des systèmes Unlock et Yahoo!PlaceMaker. Anastácio *et al.* (2010) ciblent également l'évaluation de systèmes de RENS. Les auteurs comparent les méthodes de calcul de portée spatiale supportées respectivement par les systèmes Yahoo!PlaceMaker, GIPSY, Web-a-Where et GraphRank. D'Ignazio (2013) compare les services d'analyse spatiale des systèmes de RENS OpenCalais, Clavin et Yahoo!PlaceSpotter. Enfin, Gelernter et Zhang (2013) évaluent le système de RENS Geolocator. Les auteurs mesurent la qualité du module de reconnaissance de toponymes de Geolocator ainsi que de son module d'analyse et de géolocalisation.

En ce qui concerne les systèmes de RENS, aucun *benchmark* existant ne s'est imposé comme référentiel commun. Le principal frein à l'adoption d'un *benchmark* est certainement le verrouillage (la non diffusion) des corpus et du code du logiciel d'évaluation. Nous avons pu nous procurer et travailler sur les *benchmarks* TREC-CS (Dean-Hall *et al.*, 2013) et GeoparsingQT (Gelernter, Zhang, 2013). Le premier ne propose que des ENS de type nom de grandes villes d'Amérique du nord. Le second est plus intéressant car il traite de quinze catégories différentes d'ENS. Ainsi, un jeu d'ENS hétérogènes, constitué par

des géographes, permet de mesurer les éventuels effets de bord engendrés par chaque passage à une version supérieure du système de RENS Geolocator.

Au même titre que l'évaluation de la qualité de l'annotation des EN, il est important de proposer des scénarios de visualisation des entités ainsi annotées (Luong *et al.*, 2015). La visualisation des ENS se fait dans le texte mais ces dernières peuvent également être projetées sur un fonds de carte. Il en va de même pour les entités temporelles que l'on peut représenter sur une frise du temps. Les autres entités que nous qualifions de thématiques (personnes, organisation, par exemple) peuvent être représentées sous forme de graphes ou d'arbres.

Il n'existe, à notre connaissance, qu'une plateforme intégrant des SREN. Il s'agit de la plateforme NERD [1] (*Named Entity Recognition and Disambiguation*) qui intègre les SREN Alchemy, Dandelion, DBPedia Spotlight, Lupedia, OpenCalais, Saplo, SemiTags, TextRazor, THD, Wikimeta, Yahoo!Content Analysis et Zemanta (les plus utilisés figurent dans le Tableau 1). Chacun de ces systèmes peut être invoqué via l'API de la plateforme. Les résultats sont retournés au format JSON (Rizzo *et al.*, 2012) comme illustré ci-dessous après un appel à la version 0.5 de l'API NERD :

Listing 1 – Exemple d'entité reconnue par NERD au format JSON

```
[{
        "idEntity": 35312,
        "label": "Tunisia",
        "extractorType": "Country",
        "nerdType": "http://nerd.eurecom.fr/ontology#Location",
        "uri": "http://www.cometotunisia.co.uk",
        "confidence": 0.317654,
        "relevance": 0.5,
        "extractor": "alchemyapi",
        "startChar": 1445,
        "endChar": 1452
 }]
```

Suite à cet état l'art nous soulignons la difficulté et le coût engendré par la mise en place d'une campagne d'évaluation. Ce coût est essentiellement induit par la nécessité de disposer d'experts capables, d'une part, de créer une solution de référence pour un texte donné et, d'autre part, d'évaluer la qualité des résultats retournés par les SREN afin de mesurer leur pertinence, au regard de la solution.

A ce titre, nous proposons dans la suite de cet article un modèle étendu d'EN et une plateforme permettant d'automatiser l'évaluation d'un SREN à partir d'un fichier de correction établi par un expert.

1. http://nerd.eurecom.fr/

3. SNERBM : Un cadre d'évaluation de SRENS

Nous avons repris le « benchmark » SNERBM (Palacio *et al.*, 2016), dédié à l'évaluation de la qualité des annotations spatiales, dont la démarche d'utilisation est totalement ouverte (Figure 1).

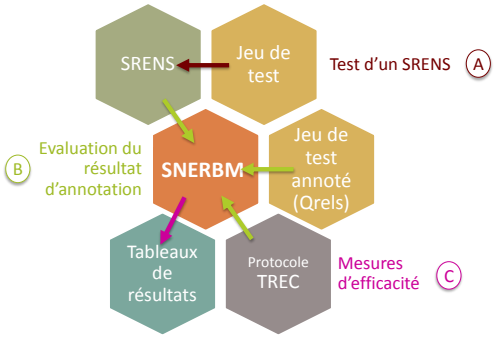

Figure 1. Le « benchmark » SNERBM

L'évalution SNERBM consiste, pour un texte donné, à comparer les entités spatiales retournées par un SRENS (A) avec les entités spatiales listées dans un fichier *Qrels* (*cf.* section 5) de correction associé à ce texte et créé par des experts. La comparaison des entités détectées par le SRENS avec les entités listées dans le fichier de correction est réalisée sur la représentation géocodée des entités.

Pour le fichier *Qrels*, les représentations géocodées sont fournies avec précision par les experts sous forme de polygones et de polylignes. Pour le fichier retourné par un SREN, deux cas de figure se présentent : soit le SRENS fournit nativement le géocodage des entités, soit ce travail est réalisé par les experts.

Le benchmark SNERBM (B) applique ensuite le protocole TREC pour la mesure d'efficacité (calcul de métriques telles que le rappel, la précision et la F1-Mesure (C) (Manning *et al.*, 2008)). Une synthèse des résultats de mesure de la qualité des annotations spatiales, produites par le SRENS sur le jeu de test, est ainsi produite.

La mise en œuvre du benchmark SNERBM nécessite l'intervention d'experts humains (Figure 2 - parties gauche et centrale) visant :

– la création du fichier de correction dans lequel chaque entité à reconnaitre est géocodée l'une après l'autre à l'aide d'un service appelé manuellement ;

– le géocodage des entités retournées par le SRENS et la mise en forme de ces résultats afin de pouvoir les comparer au fichier *Qrels* de correction ;

– l'invocation du service TREC pour comparer le fichier précédemment mis en forme avec le fichier *Qrels*.

Nous proposons une plateforme capable d'enchaîner automatiquement ces différentes étapes (Figure 2 - partie droite) de telle sorte que le géocodage des ENS reconnues et leur évaluation via le protocole TREC soit automatisés.

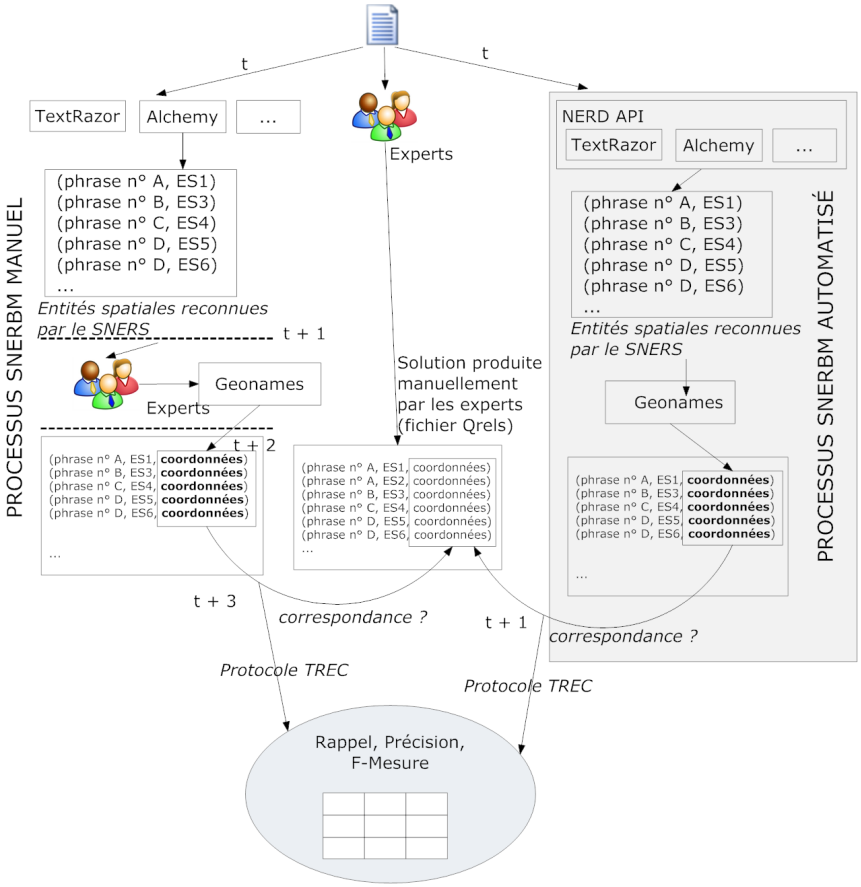

Figure 2. Automatisation du processus d'évaluation de SRENS

Dans la section suivante, nous présentons un modèle dédié à l'évaluation de SREN compatible avec le «benchmark» SNERBM.

4. Un modèle pour l'évaluation de SRENS

Conformément au cadre d'évaluation SNERBM (section 3), nous proposons un modèle permettant de faire une évaluation semi-automatique de SRENS.

Cette évaluation se base sur la comparaison des géolocalisations des entités spatiales reconnues par le SRENS (Figure 2) - il s'agit généralement de points - avec les géolocalisations des entités spatiales correspondantes dans le fichier

de correction (Fichier *Qrels* - il s'agit de géométries). Nous souhaitons réaliser cette comparaison de manière automatique en vérifiant que les géométries correspondant aux entités reconnues par le SRENS sont en intersection avec les celles relatives aux entités présentes dans le fichier de correction.

A cette fin, mais aussi à des fins de visualisation et de comparaison d'entités nommées, il est nécessaire d'étendre le modèle de données de NERD (Listing 1) repris sur la Figure 3 via la classe *Entity* (A).

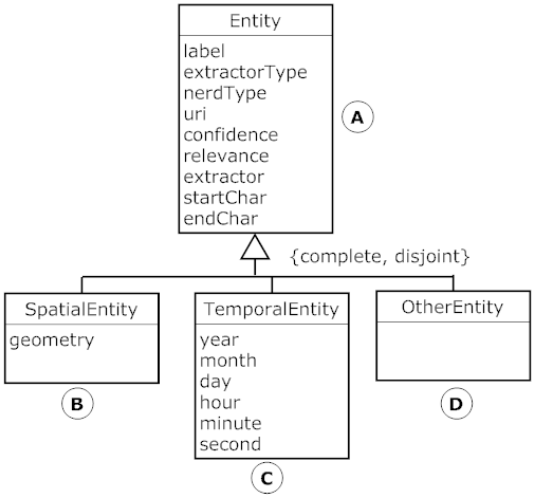

Figure 3. Modèle pivot pour l'évaluation de SRENS

Cette extension permet de :

1. catégoriser les entités reconnues en entités spatiales (B), temporelles (C) et autres (D);
2. d'enrichir chaque entité spatiale par une représentation numérique (attribut *geometry*).

Nous envisageons à terme d'appliquer la même démarche pour l'évaluation des SREN temporelles.

5. Une plateforme pour l'évaluation de SRENS dans le cadre de SNERBM

Pour mettre en place la visualisation des EN et l'évaluation des SRENS, nous proposons une plateforme modulaire présentée sur la Figure 4.

A partir d'un texte (A) sélectionné par l'utilisateur, la plateforme interroge un SRENS particulier via l'API NERD (B). Suite à cet appel, la plateforme récupère une liste d'entités nommées (C) conformément au modèle de données

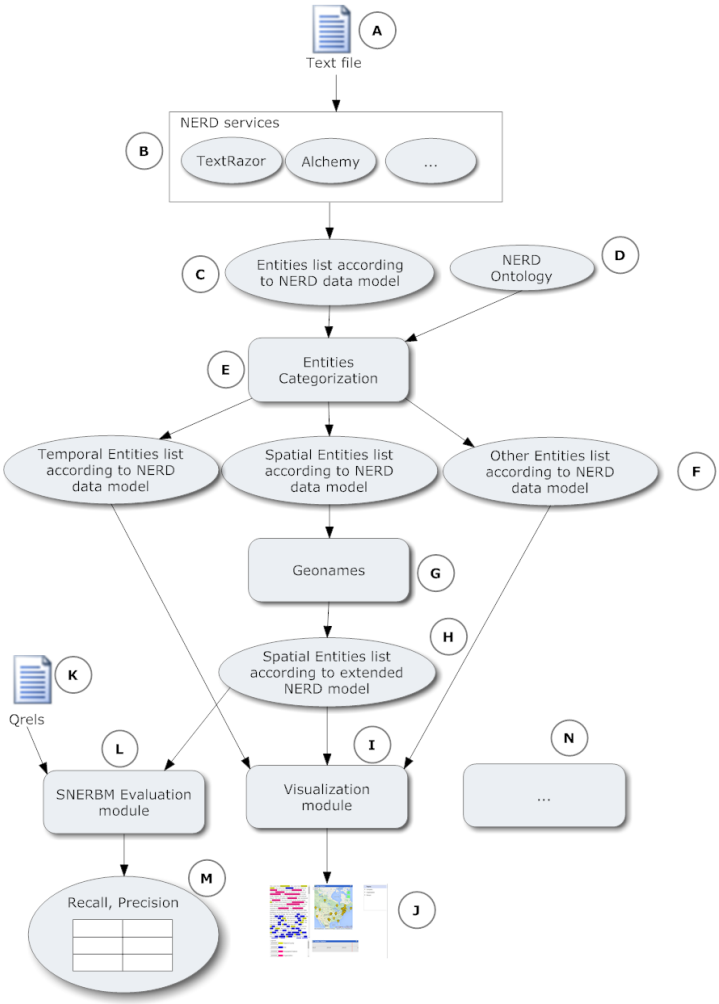

Figure 4. Architecture de la plateforme (évaluation, visualisation, fusion)

(Rizzo, Troncy, 2012) de NERD (Listing 1). En s'appuyant sur l'ontologie de NERD [2] (D), la plateforme catégorise (E) les entités retournées selon trois familles (F) : les entités spatiales, temporelles et autres.

Pour évaluer la pertinence des entités spatiales reconnues, la plateforme calcule les représentations géographiques de chaque entité via le service *Geonames* (G). Les entités spatiales reconnues par le SRENS sont donc, à ce stade,

2. http://nerd.eurecom.fr/ontology

enrichies de leur géolocalisation (H). Cette géolocalisation permet :

– d'obtenir une représentation cartographique des resultats retournés par le SRENS (J) via un module de visualisation (I);

– d'évaluer (L) automatiquement (sur la base des inclusions des points - relatifs aux entités spatiales géolocalisées - dans les polygones résultats - relatifs aux *Qrels* de correction (K)) les taux de précision et de rappel (M);

– de lancer d'autres modules permettant de (N) fusionner, désambiguiser... les entités reconnues par un ou plusieurs SRENS.

Le fichier *Qrels* de correction (K) intègre, pour chaque entité nommée à reconnaître, le numéro de phrase et la phrase où l'entité apparaît, le type de cette entité et sa représentation géocodée précise sous forme de polygone. Le code ci-dessous (Listing 2) montre un extrait de fichier *Qrels* :

Listing 2 – Extrait d'un fichier Qrels de correction

```
1 She was born in Massachusetts. ADM1     POLYGON((-73.234863 42.747012 ...))
2 She came from Belchertown.     PPL      POLYGON((-72.419643 42.300167 ...))
3 She climbed Mount Watatic.     Mountain POLYGON((-71.921310 42.708426 ...))
```

Cette plateforme, en cours de développement, propose d'ores et déjà des modules opérationnels permettant par exemple :

– de visualiser (I) les entités reconnues par un SREN (H) via une interface graphique présentant (J) les entités spatiales sur une carte, les entités temporelles sur une frise chronologique et les autres entités sous forme arborescente classées selon leur type (Figure 5) ;

– de fusionner (N) des entités spatiales reconnues par divers SRENS (H) en tenant compte du nombre (*Nbr* sur la Figure 6) et de la pertinence (R sur la Figure 6) des entités spatiales reconnues par chaque SRENS (Figure 6). Le degré de pertinence (R) est fourni par les SRENS invoqués; chaque SREN ayant sa propre échelle pour exprimer le degré de pertinence, la plateforme ramène cette valeur sur un intervalle de 0 à 1 pour homogénéiser ces résultats et les rendre comparables entre eux.

6. Expérimentation de la plateforme

Nous prenons comme cas d'application le SREN Alchemy. Il est intégré à notre plateforme et nous a permis d'expérimenter les modules d'annotation et de visualisation. La Figure 5 illustre la visualisation d'EN nommées annotées par Alchemy conformément au modèle pivot présenté en section 4. De plus, nous avons évalué l'efficacité d'Alchemy, au sein de la plateforme, en utilisant le « benchmark » SNERBM (Palacio *et al.*, 2016) sur un texte composé de 200 phrases disponibles à l'adresse suivante : http://eott.univ-pau.fr/snerbm_eval/evaluationText.txt. Nous présentons les premiers résultats d'évaluation dans le Tableau 2.

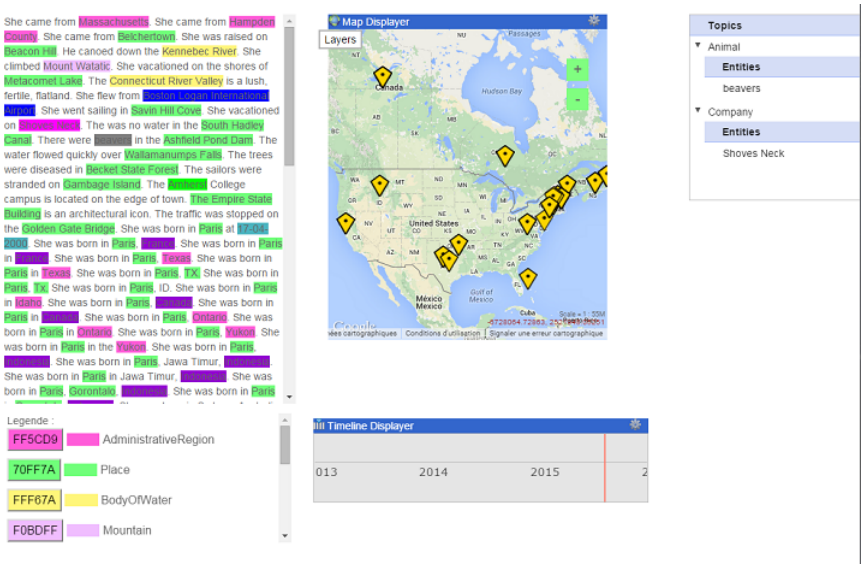

Figure 5. Illustration du module de visualisation

		PivotAlchemy.xml	PivotTextRazor.xml	fusion_07-02-2016_23-41.xml
Spatial		Nbr: 3 R:0.17	Nbr: 10 R:0	Nbr: 6 R:0.06
Country	+	Nbr: 2 R:0.17	Nbr: 6 R:0	Nbr: 2 R:0.17
Maroc		Nbr: 1 R:0.17	Nbr: 3 R:0	Nbr: 1 R:0.17
France		Nbr: 1 R:0.17	Nbr: 3 R:0	Nbr: 1 R:0.17
City	+	Nbr: 1 R:0.17	Nbr: 0 R: 0	Nbr: 0 R: 0
Amherst		Nbr: 1 R:0.17	Nbr: 0 R: 0	Nbr: 0 R: 0
University	+	Nbr: 0 R: 0	Nbr: 4 R:0	Nbr: 4 R:0
Amherst College		Nbr: 0 R: 0	Nbr: 4 R:0	Nbr: 4 R:0
Other	+	Nbr: 0 R: 0	Nbr: 0 R: 0	Nbr: 0 R: 0
Temporal	+	Nbr: 0 R: 0	Nbr: 0 R: 0	Nbr: 0 R: 0

Figure 6. Fusion d'entités spatiales reconnues par plusieurs SRENS

Tableau 2. Evaluation d'Alchemy via la plateforme et le cadre SNERBM

Outil de REN	Alchemy
Nombre total d'ENS	317
Nombre total d'ENS reconnues	197
Nombre total d'ENS pertinentes	93
Rappel	0.47
Précision	0.29
F1-Mesure	0.35

Le rappel (pourcentage d'annotations spatiales pertinentes par rapport au nombre total d'ENS dans le document) est de 47%. La précision (pourcentage d'annotations spatiales pertinentes par rapport au nombre total d'annotations spatiales) est de 29%. Enfin, la F1-Mesure correspondante est de 35%. Ces premiers résultats sont encourageants et nous allons procéder à l'évaluation d'autres SRENS intégrés dans la plateforme.

Nous envisageons par la suite d'exploiter les indicateurs de rappel, précision et F1-Mesure dans le module de fusion présenté Figure 6. Ce module permet de fusionner des entités reconnues par différents SREN sur un même texte. Les critères de fusion proposés se basent actuellement uniquement sur le nombre (pour privilégier la quantité) ou la pertinence (pour privilégier la qualité) des entités reconnues. Il sera intéressant d'étendre ces indicateurs en intégrant les retours proposés par le module d'évaluation pour, par exemple, fusionner des entités en fonctions des valeurs de précision, rappel ou F1-Mesure calculées.

7. Conclusion et perspectives

Dans cet article, nous avons proposé une plateforme et un modèle dédiés à l'évaluation semi-automatique de SRENS conformément au cadre d'évaluation SNERBM. Le modèle proposé est une extension du modèle NERD permettant d'homogénéiser les retours des SRENS. L'extension consiste à enrichir automatiquement la représentation des entités spatiales reconnues en leur adjoignant une représentation numérique sous forme de coordonnées géographiques. L'ajout de ces coordonnées permet entre autres :

– d'identifier de manière unique chaque entité reconnue,

– de comparer automatiquement les entités reconnues par le SREN avec les entités listées dans la solution des experts,

– de représenter chaque entité sur un support cartographique afin de faciliter une évaluation visuelle des résultats retournés par le SRENS.

Cette extension du modèle NERD apporte une plus-value pour mettre en oeuvre le benchmark SNERBM car elle permet d'automatiser le processus d'évaluation. Les experts doivent toujours produire manuellement un fichier de correction type pour un texte donné mais une fois ces fichiers créés le processus d'évaluation peut être lancé automatiquement pour obtenir directement une évaluation selon des critères de rappel, précision et F1-Mesure. Il devient alors peu couteux de lancer une série d'évaluations impliquant différents SRENS afin de les situer les uns par rapport aux autres.

Nous envisageons plusieurs perspectives à court terme concernant l'évaluation des SRENS :

– Désambiguiser les retours de la plateforme Geonames qui propose parfois plusieurs géocodages pour une même ENS (par défaut, la plateforme conserve uniquement le premier géocodage retourné) ;

– Intégrer dans le fichier de correction les types d'entités spatiales à reconnaitre (commune, pays, fleuve ...) et prendre en compte une correspondance des types lors de la comparaison ;

– Mesurer l'éventuelle perte d'efficacité des SREN due à leur intégration dans une plateforme qui impose la transformation des annotations dans un modèle pivot générique (Figure 3). Pour cela, nous envisageons d'évaluer deux versions de chaque SRENS selon le « benchmark » SNERBM : (i) le SRENS est invoqué directement pour annoter le jeu de données ; (ii) le SRENS est invoqué via notre plateforme support (avec usage du modèle pivot) pour annoter le même jeu de données.

A moyen terme, nous envisageons d'étendre ce processus d'évaluation aux entités temporelles en appliquant ce principe d'intersection sur des données temporelles formalisées sous forme calendaire telle que déjà définie dans le modèle pivot présenté Figure 3 (C).

A plus long terme nous visons à étendre la plateforme d'évaluation pour intégrer des modules permettant entre autres :

– d'intégrer d'autres SRENS que ceux proposés par l'API NERD ;

– d'ajouter un module permettant de désambiguiser les propositions de certains SREN qui retournent plusieurs entités candidates pour une même annotation textuelle.

Bibliographie

Anastácio I., Martins B., Calado P. (2010). Using the geographic scopes of web documents for contextual advertising. In *GIR'10: Proceedings of the 6th Workshop on Geographic Information Retrieval*, p. 18:1–18:8. New York, NY, USA, ACM.

Andogah G. (2010). *Geographically Constrained Information Retrieval.* Thèse de doctorat, University of Groningen, Netherlands.

Bucher B., Clough P., Joho H., Purves R., Syed A. K. (2005). Geographic IR Systems: Requirements and Evaluation. In *ICC'05: Proceedings of the 22nd International Cartographic Conference.* Global Congressos.

Chinchor N. A. (1998). MUC/MET evaluation trends. In *Proceedings of the TIPSTER text program: Phase III*, p. 235–239. Baltimore, Maryland, USA, Association for Computational Linguistics.

Chinchor N. A. (1998). Overview of MUC-7. In *MUC-7: Proceedings of the 7th Message Understanding Conference.*

Dean-Hall A., Clarke C. L. A., Kamps J., Thomas P., Simone N., Voorhees E. (2013). Overview of the TREC 2013 Contextual Suggestion Track. In *TREC'13: Proceedings of the 22nd text retrieval conference.* Gaithersburg, MD, NIST.

D'Ignazio C. (2013). *Big data, news and geography: Research update.* Consulté sur https://civic.mit.edu/blog/kanarinka/big-data-news-and-geography (MIT Center for Civic Media)

Gelernter J., Zhang W. (2013). Cross-lingual geo-parsing for non-structured data. In *GIR'13: Proceedings of the 7th Workshop on Geographic Information Retrieval*, p. 64–71. New York, NY, USA, ACM.

Grishman R., Sundheim B. (1996). Message understanding conference- 6: A brief history. In *16th international conference on computational linguistics, proceedings of the conference, COLING 1996, center for sprogteknologi, copenhagen, denmark, august 5-9, 1996*, p. 466–471.

Leidner J. L. (2007). *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names.* Thèse de doctorat, Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh, Scotland.

Lenzi V. B., Speranza M., Sprugnoli R. (2013). Named entity recognition on transcribed broadcast news at EVALITA 2011. In *Revised Papers from EVALITA'11: International Workshop on the Evaluation of Natural Language and Speech Tools for Italian*, vol. 7689, p. 86-97. Springer.

Lieberman M. D., Samet H., Sankaranarayanan J. (2010). Geotagging with local lexicons to build indexes for textually-specified spatial data. In F. Li *et al.* (Eds.), *Proceedings of the 26th international conference on data engineering, ICDE 2010, march 1-6, 2010, long beach, california, USA*, p. 201–212. IEEE.

Lingad J., Karimi S., Yin J. (2013). Location extraction from disaster-related microblogs. In L. Carr *et al.* (Eds.), *22nd international world wide web conference, WWW '13, rio de janeiro, brazil, may 13-17, 2013, companion volume*, p. 1017–1020. International World Wide Web Conferences Steering Committee / ACM.

Luong T. N., Marquesuzaà C., Etcheverry P., Nodenot T., Laborie S. (2015). Facilitating the design/evaluation process of web-based geographic applications: A case study with windmash. In T. K. Dang, R. Wagner, J. Küng, N. Thoai, M. Takizawa, E. J. Neuhold (Eds.), *Future data and security engineering - second international conference, FDSE 2015, ho chi minh city, vietnam, november 23-25, 2015, proceedings*, vol. 9446, p. 259–271. Springer.

Mandl T., Carvalho P., Nunzio G. M. D., Gey F. C., Larson R. R., Santos D. *et al.* (2009). GeoCLEF 2008: The CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview. In C. Peters *et al.* (Eds.), *Revised Selected Papers of CLEF'08: Proceedings of the 9th Workshop of the Cross-Language Evaluation Forum*, vol. 5706, p. 808–821.

Manning C. D., Raghavan P., Schütze H. (2008). *Introduction to Information Retrieval.* Cambridge University Press.

Marrero M., Sánchez-Cuadrado S., Lara J. M., Andreadakis G. (2009). Evaluation of named entity extraction systems. *Advances in Computational Linguistics. Research in Computing Science*, vol. 41, p. 47–58.

Nouvel D. (2012). *Reconnaissance des entités nommées par exploration de règles d'annotation.* Thèse de doctorat, Université François Rabelais de Tours, France.

Palacio D., Sallaberry C., Cabanac G., Hubert G. (2016). Cadre d'évaluation de systèmes de reconnaissance d'entités nommées spatiales. In *Actes du XXXIVe congrès INFORSID, Grenoble, France, may 31 - june 3, 2016.*, p. 149–164. Consulté sur http://inforsid.fr/actes/2016/INFORSID2016_paper_9.pdf

Rizzo G., Troncy R. (2012, 04/2012). Nerd: A framework for unifying named entity recognition and disambiguation web extraction tools. In *European chapter of the association for computational linguistics (EACL'12).* Avignon, France.

Rizzo G., Troncy R., Hellmann S., Brümmer M. (2012). NERD meets NIF: lifting NLP extraction results to the linked data cloud. In C. Bizer, T. Heath, T. Berners-Lee, M. Hausenblas (Eds.), *WWW2012 workshop on linked data on the web, lyon, france, 16 april, 2012*, vol. 937.

Santos D., Seco N., Cardoso N., Vilela R. (2006). HAREM: An Advanced NER Evaluation Contest for Portuguese. In *LREC'06: Proceedings of the 5th International Conference on Language Resources and Evaluation*, p. 1986–1991.

Sekine S., Eriguchi Y. (2000). Japanese Named Entity Extraction Evaluation – Analysis of Results. In *COLING'00: Proceedings of the 18th conference on Computational linguistics*, p. 1106-1110. Association for Computational Linguistics.

Strassel S., Przybocki M. A., Peterson K., Song Z., Maeda K. (2008). Linguistic Resources and Evaluation Techniques for Evaluation of Cross-Document Automatic Content Extraction. In *LREC'08: Proceedings of the International Conference on Language Resources and Evaluation*, p. 2706–2709.

Tjong Kim Sang E. F., De Meulder F. (2003). Introduction to the CoNLL-2003 Shared Task Language-Independent Named Entity Recognition. In W. Daelemans, M. Osborne (Eds.), *CoNLL-2003: Proceedings of the 7th conference on Natural language learning at HLT-NAACL 2003*, p. 142–147. Association for Computational Linguistics.

Tobin R., Grover C., Byrne K., Reid J., Walsh J. (2010). Evaluation of georeferencing. In R. Purves, P. Clough, C. B. Jones (Eds.), *GIR'10: Proceedings of the 6th Workshop on Geographic Information Retrieval.* New York, NY, USA, ACM.

Voorhees E. M. (2001). Overview of TREC 2001. In *TREC'01: Proceedings of the 9th Text REtrieval Conference.* Gaithersburg, MD, USA.