Vers la construction d'une base de connaissances sur la réorganisation territoriale française à la Révolution

Antoine Keller, Nathalie Abadie, Bertrand Dumenieu, Stéphane Baciocchi, Eric Kergosien

Vers la construction d'une base de connaissances sur la réorganisation territoriale française à la Révolution

Antoine Keller[1], **Nathalie Abadie**[2], **Bertrand Duménieu**[3], **Stéphane Baciocchi**[3], **Eric Kergosien**[4]

1. *École Navale, Lanvéoc*
   keller.a.pro@gmail.com
2. *Univ. Paris-Est, LASTIG COGIT, IGN, ENSG, Saint-Mandé*
   nathalie-f.abadie@ign.fr
3. *CRH, EHESS, Paris*
   bertrand.dumenieu@ehess.fr;stephane.baciocchi@ehess.fr
4. *GERiiCO, Univ. Lille 3, Villeneuve-d'Ascq, Lille*
   eric.kergosien@univ-lille3.fr

RÉSUMÉ. *Cet article présente une approche pour créer une base de connaissances spatio-temporelles sur la réorganisation du découpage administratif français lors de la Révolution française (1790-1793). Cette base est construite en associant les informations spatiales de la carte de Cassini et les décrets de l'Assemblée Nationale décrivant finement les remembrements paroissiaux. L'approche que nous proposons consiste à reconnaître et désambiguïser les Entités Spatiales Nommées citées dans les décrets à l'aide d'un gazetier issu de la carte de Cassini, les relations spatio-temporelles dues à la réorganisation des paroisses et structurer ces informations dans un graphe.*

ABSTRACT. *In this article we present an approach to create a spatio-temporal knowledge base on the reorganization of the French parishes during Revolution (1790-1793). This database is built on the combination of spatial information coming from the Cassini map and the decrees of the National Constituen Assembly. The approach consists in recognizing and disambiguating the Spatial Named Entities cited in the decrees and the spatio-temporal relationships specific to the reorganization of parishes. Extracted information is then structured into a RDF graph.*

MOTS-CLÉS : *Extraction d'information à partir de textes, base de connaissances, données géohistoriques, évolutions du territoire*

KEYWORDS: *Text mining, knowledge base, geohistorical data, spatial dynamics*

1. Introduction

L'essor des humanités numériques a ouvert la voie à la publication et la réutilisation massive de données historiques. Parmi elles, les ressources cartographiques et iconographique sont de plus en plus régulièrement exploitées dans des Systèmes d'Information Géographique (SIG) pour servir de référentiels spatiaux, supports d'études sur la connaissance et l'évolution historiques du territoire (Biszak *et al.*, s. d.; Costes, 2016; Dumenieu, 2015; Cura *et al.*, 2018). Les archives textuelles, imprimés ou manuscrits sont aussi une formidable source d'informations sur l'espace ancien et ses évolutions, mais elles sont souvent négligées ou sous-utilisées car difficiles à extraire et structurer. Ainsi, le remembrement des paroisses religieuses d'Ancien Régime lors de la mise en place de l'Église constitutionnelle (1790-1793) est très précisément décrit dans les archives des débats et délibérations de l'Assemblée Constituante. Ces textes permettent de reconstituer finement la réorganisation du maillage paroissial. Une base de connaissances spatialisée contenant les informations de ces décrets rendrait possible l'analyse quantitative d'un aspect de la réorganisation du territoire français à la Révolution et une première cartographie systématique des limites paroissiales de la France d'Ancien Régime, encore aujourd'hui mal connues. Les premières étapes de la construction d'une telle base constituent l'objet de cet article.

Après un état de l'art des approches pour l'extraction d'entités spatiale nommées à partir de textes, leur désambiguïsation, et leur structuration, nous présentons les sources d'information dont nous disposons pour reconstituer le remembrement paroissial dû à la mise en place de l'Église constitutionnelle. Le reste de l'article décrit l'approche envisagée pour créer une base de connaissances spatio-temporelles sur ce phénomène en vue d'analyses futures.

2. Etat de l'art

Dans le sillage des travaux en extraction d'entités nommées encouragés par les conférences MUC [1] (Chinchor, 1998), la reconnaissance et la désambiguïsation d'entités spatiales nommées ont fait l'objet de nombreuses propositions au cours des dernières années, pouvant être mis à profit pour l'extraction de connaissances dans des corpus géohistoriques.

2.1. Extraction d'entités spatiales à partir de textes

Lesbegueries (2007) distingue les entités spatiales nommées absolues, qui correspondent à des noms de lieux éventuellement accompagnés d'un concept

1. MUC désigne les conférences à l'initiative et financées par la DARPA ayant pour but l'amélioration des techniques d'extraction d'information.

topographique, comme *la ville de Paris*, des entités spatiales relatives, nommées ou pas, qui désignent des entités spatiales définies par rapport à une ou plusieurs autres entités spatiales absolues à l'aide d'une ou plusieurs relations spatiales, comme *le fleuve qui traverse Paris*.
Les approches de reconnaissance et de désambiguïsation d'entités spatiales nommées (ESN) spécialisent celles destinées aux entités nommées (EN) pour prendre en compte la nature géographique des entités à traiter. Une approche classique de traitement du langage naturel pour la reconnaissance d'EN comporte quatre étapes : (i) tokenisation du corpus (ii) analyse morpho-lexicale (iii) analyse syntaxique (iv) analyse sémantique (Abolhassani *et al.*, 2003). La tokenisation consiste à découper le texte en *tokens*, unités lexicales qui pourront par la suite être analysées individuellement. L'analyse morpho-lexicale s'appuie sur des ressources lexicales pour retrouver la forme canonique des mots (leur lemme) et identifier leur nature et leur forme grammaticale. L'analyse syntaxique identifie les relations entre mots au sein d'une phrase en commençant par rechercher la fonction grammaticale des mots qui la composent. Enfin l'analyse sémantique cherche à donner du sens à la phrase; on peut par exemple utiliser des patrons pour identifier des mots présentant un intérêt particulier et analyser leur contexte. Cette chaîne de traitement vise à identifier ou produire des descripteurs (Nadeau, Sekine, 2007) qui sont ensuite utilisés pour reconnaître des ESN et éventuellement leurs relations à l'aide de règles définies manuellement, de modèles appris par des algorithmes de classification automatique ou encore d'un combinaison de ces deux types d'approches (Wu *et al.*, 2006). Les approches à bases de règles reposent sur des grammaires locales définies manuellement et de façon empirique par un expert en sciences du langage. Les approches proposées par (Stern, Sagot, 2010; Moncla, 2015) combinent l'application de gazetiers, index composés de paires associant un toponyme à des coordonnées géographiques, et la recherche de termes discriminants dans le contexte des noms propres identifiés dans les textes comme "ville", "col", ou "route". Les approches par apprentissage supervisé consistent à entraîner le système avec un corpus où les E(S)N sont déjà annotées pour apprendre un modèle fondé sur des descripteurs de mots préalablement choisis pour leur caractère discriminant. Ce modèle est appliqué au corpus à traiter pour classer les mots selon leur appartenance à une catégorie d'EN (Finkel *et al.*, 2005; Raymond, Fayolle, 2010; Tkachenko, Simanovsky, 2012). Elles permettent de s'affranchir de l'étape de création manuelle de règles mais nécessitent un corpus d'entraînement conséquent.

2.2. Approches pour la désambiguïsation d'ESN

L'intégration des entités spatiales extraites des textes nécessite leur désambiguïsation, c'est-à-dire l'identification du lieu qu'elles désignent au sein d'un référentiel géographique. Pour ce faire, les ESN sont comparées aux noms de lieux fournis par le référentiel géographique à l'aide d'une mesure de similarité

de chaînes de caractères. Les noms de lieux les plus similaires sont alors retenus comme candidats.
Pour les départager Buscaldi (2011) distingue trois types d'approches. Les approches fondées sur les coordonnées suivent l'intuition selon laquelle des ESN mentionnées dans une même portion de texte sont susceptibles de désigner des lieux proches (Derungs, Purves, 2014; Zhao *et al.*, 2014). Les approches de ce type s'appuient sur les ESN non ambiguës et calculent leur distance par rapport aux ESN ambiguës, les plus proches étant alors retenus. D'autres approches, pilotées par les données, utilisent des algorithmes d'apprentissage supervisé sur diverses propriétés descriptives des lieux nommés et les classent en fonction de ces propriétés (Buscaldi, 2011; Santos *et al.*, 2015). Enfin, les approches fondées sur les connaissances s'appuient sur d'autres informations extraites du corpus (EN, relations entre EN, etc.) et sur des bases de connaissances externes comme DBpedia, Yago ou Wordnet pour classer les candidats. Lorsqu'une ESN a plusieurs candidats dans la base de connaissances, les informations additionnelles extraites du texte sont comparées aux connaissances liées à chacun des candidats dans la base de connaissances pour estimer le degré de similarité entre ces informations (Ireson, Ciravegna, 2010; Speriosu, Baldridge, 2013; Batista *et al.*, 2012). En l'absence de référentiel géographique doté de coordonnées, une alternative consiste donc à substituer au calcul de distance l'évaluation de la proximité des candidats avec l'ESN non ambiguë au sein de la hiérarchie d'un gazetier (Amitay *et al.*, 2004). Une variante de ces approches consiste à mettre à profit la structure de graphe de la base de connaissances pour classer les candidats (Brando *et al.*, 2016; Paris *et al.*, 2017; Moro *et al.*, 2014).

2.3. Structuration de connaissances spatio-temporelles

Les informations disponibles et pouvant être extraites automatiquement à partir des textes sont à la fois très riches et partielles. Elles sont donc difficilement intégrables dans les modèles de données des SIG qui requièrent des données fortement structurées a priori (OGC, 2011). En particulier, ceux-ci requièrent des données dotées de références spatiales directes (coordonnées ou géométries) qui ne sont pas nécessairement disponibles dans le cas d'entités spatiales nommées extraites de textes. Cette contrainte a notamment incité les créateurs du gazetier géohistorique Pleiades à s'orienter vers le modèle de données en graphe pour le Web sémantique : RDF [2] (Elliott, Gillies, 2008).
(Chen *et al.*, 2018) proposent un modèle de base de données en graphe pour structurer et stocker des connaissances extraites à partir de descriptions textuelles de lieux. Ce modèle étend un modèle de graphe initial (Kim *et al.*, 2015) fondé sur la mise en commun de triplets créés par extraction d'expressions décrivant de façon directe ou indirecte la localisation d'ES sous forme textuelle. Dans ce modèle, l'étape de désambiguïsation des ES est cruciale pour permettre

2. https://www.w3.org/RDF/

une bonne intégration des triplets. En effet, elle permet d'identifier les mentions d'ES faisant référence à un même lieu et donc de lier les triplets créés à partir de ces mentions.
Diverses approches ont été proposées ces dernières années pour constituer de grandes bases de connaissances sur le Web de données, qu'il s'agisse d'approches collaboratives (Vrandečić, Krötzsch, 2014; Bollacker *et al.*, 2008) ou d'approches par extraction d'informations dans des contenus du Web. Ainsi, DBPedia (Lehmann *et al.*, 2015) et Yago (Rebele *et al.*, 2016), extraites automatiquement à partir des contenus de Wikipedia, constituent les exemples les plus connus de ce second type d'approches. Dans ces deux exemples l'approche repose très largement sur les contenus structurés de l'encyclopédie libre (infoboxes, templates, catégories et autres liens), plus faciles à extraire et intégrer (Weikum *et al.*, 2016). Dans tous les cas, les standards du Web sémantique sont largement mis à profit pour représenter les connaissances extraites et pour en évaluer et améliorer la qualité (Zaveri *et al.*, 2016; Paulheim, 2017).

3. Corpus et données topographiques de référence utilisés

3.1. Les décrets de circonscription des paroisses constitutionnelles

L'Église romaine condamnant la Révolution, une Constitution Civile du Clergé fut mise en place afin de réorganiser les rapports entre les prêtres, les paroisses et la monarchie constitutionnelle française. Cela se concrétisa notamment par une redéfinition territoriale des évêchés, correspondant désormais aux nouveaux départements, et des paroisses, comprises dans les limites communales. Le déclassement d'une cinquantaine de sièges épiscopaux et la suppression de « plusieurs centaines de paroisses » ont ainsi démultiplié les fermetures d'églises [3], notamment dans les villes où, suivant le nouvel ordre géométrique et égalitaire des choses, le nombre de paroisses et de succursales se trouvait généralement « disproportionné » par rapport à celui de la population résidente (Baciocchi, Julia (2009)). Les cités épiscopales et les grandes villes riches en

3. Un dépouillement systématique des décrets de circonscription paroissiale entérinés par l'Assemblée nationale fait apparaître, en première approximation, que jusqu'au mois de septembre 1791, et suivant un rythme qui s'est accéléré au printemps 1791, l'ensemble des territoires ayant fait l'objet d'une procédure aboutie de circonscription constitutionnelle a perdu environ 60% de ses paroisses d'Ancien Régime. Cette réduction porta en priorité sur les régions Nord, Île-de-France et Centre, et toucha tout particulièrement les cités épiscopales et les localités de moins de 6 000 habitants dont les paroisses urbaines furent systématiquement réduites à une seule. Voir A. P., t. 20, p. 351-352 ; t. 22, p. 101-102, 422, 476-477, 516-517, 739-744 ; t. 23, p. 112, 172, 221-223, 651, 657-658 ; t. 24, p. 31, 86-87, 143-144, 292-293, 493-494, 559-560, 578 ; t. 25, p. 1-2, 232-233, 235, 326-327, 375-376, 412-413, 432, 553, 555-556, 575-576, 864 ; t. 26, p. 29-30, 575-576, 694-697 ; t. 27, p. 139-140, 189, 251-253, 759-763 ; t. 28, p. 595-596 ; t. 29, p. 259-260, 472-477, 641-644 ; t. 30, p. 25-26, 93, 303-304, 559-561, 627-629 ; t. 31, p. 123-129.

sanctuaires, tout comme les petites villes et les bourgs de moins de six mille habitants obligés de réduire à une seule leurs paroisses, furent tout particulièrement affectés par ces mesures législatives. Les décrets de circonscription des paroisses sont disponibles dans les fonds des archives parlementaires et présentent un intérêt particulier dans la mesure où ils s'appuient, pour la remembrer, sur l'organisation paroissiale d'Ancien Régime. Or, celle-ci est mal connue : on la devine, sans réellement en connaître les contours, à partir des clochers représentés sur les cartes des Cassini. De plus, ces décrets donnent un premier aperçu sur les liens entre paroisses et communes. En effet, les paroisses d'Ancien Régime supprimées, les registres paroissiaux désormais sécularisés (on parle depuis d'état-civil) passèrent aux chefs-lieux des paroisses constitutionnelles. Dans un premier temps, on travaillera sur les textes de quatre séances de l'Assemblée Nationale qui comprennent [4], entre autres, la circonscriptions des paroisses du Puy-de-Dôme. Ceci forme un corpus de 20 pages d'archives, numérisées et dont le contenu a été extrait par OCR.

3.2. Une base de données géographique sur la France du XVIII^{e}s.

L'identification et la désambiguïsation des entités spatiales nommées s'appuie traditionnellement sur des gazetiers. L'utilisation de gazetiers modernes pour identifier des lieux du XVIII^{e} siècle est problématique car anachronique. En 230 ans, les changements toponymique alliés aux remodelages continuels du territoire sont suffisamment importants pour les disqualifier. Dans l'objectif de fournir une description structurée du territoire français à l'époque moderne, le groupe de recherche pluridisciplinaire GeoHistoricalData produit collaborativement un jeu de données géographiques historiques (géohistoriques) à partir de la carte de Cassini. La *Carte Générale de la France*, dite "de Cassini" est le produit d'une des principales aventures scientifiques et techniques du XVIII^{e} siècle : levée géométriquement [5] sous la direction de César-François Cassini à partir de 1756 et terminée d'imprimer au moment de la Révolution, elle dessine au 1 : 86400 le territoire du Royaume de France de la dernière moitié du siècle. Après géoréférencement, une partie des thèmes de la carte a été vectorisée par les membres de l'équipe GeoHistoricalData, mais seul le réseau routier et l'ensemble des chefs-lieux sont à ce jour complets. Parmi eux, celui regroupant les lieux nommés nous intéresse ici en particulier. La carte les localise au moyen d'un ensemble de symboles graphiques, chaque symbole correspondant à un type de lieu (on parlerait aujourd'hui de *featuretype*). Ce jeu de données géohistorique associe aux géométries ponctuelles leur dénomination ainsi que leur type cartographique, c'est à dire le symbole qui le représente sur la carte. Ainsi,

4. Séances du 1er juin 1791 au soir, du 15 juin 1791 au soir, 16 août 1791 au soir et 21 septembre 1791.

5. Le tracé de la carte s'appuie sur un canevas de triangles couvrant la France, lui-même construit en plusieurs phases à partir de 1730 (Thury, 1783; Puissant, 1812; Gallois, 1909; Konvitz, 1982)

la quasi-totalité des lieux de type "clocher" correspondent à des chefs lieux de paroisses d'Ancien Régime.

4. Création d'une base de connaissances spatio-temporelle à partir des décrets

Ce travail vise à construire une base de connaissances structurée sous forme de graphe spatio-temporel *linked-data* exploitant le modèle Snapshot (Peuquet, 1994), proche de propositions existantes (Del Mondo G, Stell J G, Claramunt C, et Thibaud R, 2010; Harbelot *et al.*, 2015). Pour cela, nous proposons une approche s'appuyant sur l'état de l'art pour extraire automatiquement un tel graphe à partir des décrets de circonscription des paroisses. Celle-ci soulève des questions de recherche dans au moins trois domaines (Chen *et al.*, 2018) : l'extraction d'informations à partir de textes, la modélisation d'informations et la gestion de connaissances.

La figure 1 présente un graphe qui pourrait être construit à partir de l'extrait suivant : *"Bourg-Lasticq, à laquelle sera réunie la paroisse de Saint-Sulpice, distraction faite des hameaux de Lasticq, Méauzat et Granges.[...] Herment, qui comprendra, outre son territoire, les hameaux de Lasticq, Méauzat et Granges, distraits de Saint-Sulpice; et le hameau de Laveix, les domaines de la Conche, Barberolles et Villevault, distraits de la paroisse de Verneugeol.[...] Verneugeol, qui n'éprouvera d'autre changement que la distraction faite d'une partie de son territoire, en faveur d'Herment."* Il se compose de deux *snapshots* spatio-temporels représentant les états de l'organisation paroissiale respectivement avant et après la réorganisation, les relations se répartissent en deux catégories. D'une part, des relations hiérarchiques entre lieux et d'autre part les relations spatio-temporelles de continuation ("sameas") ou dérivation ("devient"). La chaîne de traitement envisagée pour construire une telle base de connaissance est illustrée par la figure 2. Elle comporte quatre étapes : (1) la reconnaissance des ESN dans les textes des décrets, (2) la désambiguïsation des ESN à l'aide de la base de données des lieux ponctuels de la carte de Cassini, (3) la reconnaissance de relations entre ESN et enfin (4) la structuration des informations extraites en graphe RDF et la vérification de sa cohérence.

Les ESN à extraire du corpus ayant de grandes chances de figurer dans la base de données des lieux ponctuels de Cassini, une approche à base de règles mettant à profit cette ressource semble constituer un choix raisonnable. En outre, la reconnaissance par apprentissage supervisé paraît peu pertinente pour ce premier travail dans la mesure où l'échantillon de décrets est de petite taille. Cette étape s'inspire de l'approche proposée par Moncla (2015) qu'elle adapte en mettant en oeuvre un gazetier géohistorique et un vocabulaire adapté aux types d'entités extraites. S'agissant de textes de loi, les formulations utilisées pour décrire les évolutions des paroisses d'ancien régime sont relativement précises et permettent également d'envisager le recours à des patrons pour la

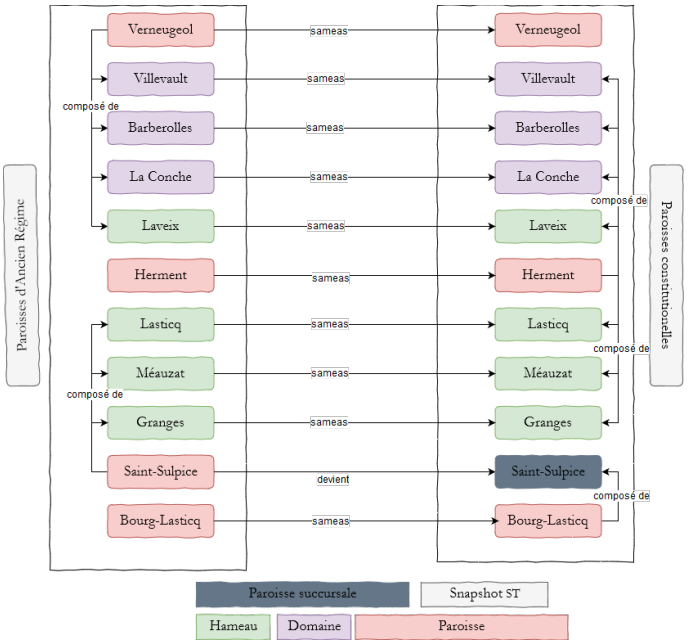

FIGURE 1 – Un exemple de représentation de la réorganisation de quelques paroisses sous forme de graphe spatio-temporel.

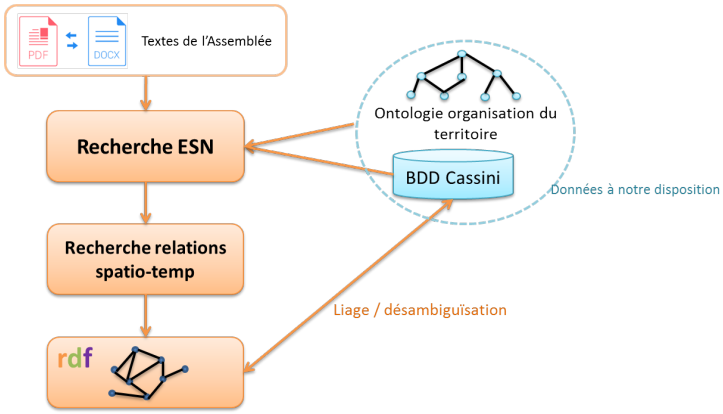

FIGURE 2 – Chaîne de traitement

reconnaissance des relations entre ESN. Enfin, l'objet même de ces décrets permet d'espérer de bons résultats en adoptant une approche de désambiguïsation fondée sur les coordonnées (Moncla *et al.*, 2014).

5. Conclusion

Cette article propose une approche de construction d'une base de connaissances spatio-temporelle sur la réorganisation du tissu paroissial français pendant la Révolution, à partir de deux sources historiques majeures : les décrets de l'Assemblée Constituante et la carte de Cassini. L'approche présentée ici est en cours de développement. L'étape suivante consistera à identifier les relations hiérarchiques et de transformations qui lient les entités spatiales nommées citées dans les décrets.

Bibliographie

Abolhassani M., Fuhr N., Gövert N. (2003). Information extraction and automatic markup for xml documents. , p. 159–174.

Amitay E., Har'El N., Sivan R., Soffer A. (2004). Web-a-where: geotagging web content. In *Proceedings of the 27th annual international acm sigir conference on research and development in information retrieval*, p. 273–280.

Baciocchi S., Julia D. (2009). Reliques et révolution française (1789-1804). *Reliques modernes: cultes et usages chrétiens des corps des saints des Réformes aux revolutions*, vol. 2, p. 483–585.

Batista D. S., Ferreira J. D., Couto F. M., Silva M. J. (2012). Toponym disambiguation using ontology-based semantic similarity. In *International conference on computational processing of the portuguese language*, p. 179–185.

Biszak E., Biszak S., Timár G., Nagy D., Molnár G. (s. d.). Historical topographic and cadastral maps of europe in spotlight–evolution of the mapire map portal.

Bollacker K., Evans C., Paritosh P., Sturge T., Taylor J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 acm sigmod international conference on management of data*, p. 1247–1250.

Brando C., Frontini F., Ganascia J.-G. (2016). Reden: named entity linking in digital literary editions using linked data sets. *Complex Systems Informatics and Modeling Quarterly*, n° 7, p. 60–80.

Buscaldi D. (2011). Approaches to disambiguating toponyms. *Sigspatial Special*, vol. 3, n° 2, p. 16–19.

Chen H., Vasardani M., Winter S., Tomko M. (2018). A graph database model for knowledge extracted from place descriptions. *ISPRS International Journal of Geo-Information*, vol. 7, n° 6. Consulté sur http://www.mdpi.com/2220-9964/7/6/221

Chinchor N. (1998). Overview of muc-7. In *Seventh message understanding conference (muc-7): Proceedings of a conference held in fairfax, virginia, april 29-may 1, 1998.*

Costes B. (2016). *Vers la construction d'un référentiel géographique ancien: un modèle de graphe agrégé pour intégrer, qualifier et analyser des réseaux géohistoriques.* Thèse de doctorat non publiée, Paris Est.

Cura R., Dumenieu B., Abadie N., Costes B., Perret J., Gribaudi M. (2018). Historical collaborative geocoding. *ISPRS International Journal of Geo-Information*, vol. 7, n° 7, p. 262.

Del Mondo G, Stell J G, Claramunt C, et Thibaud R. (2010). A graph model for spatio-temporal evolution. *Journal of Universal Computer Science*, vol. 16, n° 10, p. 1452–1477.

Derungs C., Purves R. S. (2014). From text to landscape: locating, identifying and mapping the use of landscape features in a swiss alpine corpus. *International Journal of Geographical Information Science*, vol. 28, n° 6, p. 1272–1293.

Dumenieu B. (2015). *Un système d'information géographique pour le suivi d'objets historiques urbaines à travers l'espace et le temps.* Thèse de doctorat non publiée, EHESS. Consulté sur http://bibliosr.ign.fr/Publications/2015/Dumenieu15 (Spécialité mathématiques et applications aux sciences de l'homme.)

Elliott T., Gillies S. (2008). Pleiades: the un-gis for ancient geography. *Journal of Geographical Information Science*, vol. 22, p. 1091–1108.

Finkel J. R., Grenager T., Manning C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, p. 363–370.

Gallois L. (1909). L'académie des sciences et les origines de la carte de cassini. *Annales de Géographie*, vol. 18, n° 100, p. 193–204 et 289–310.

Harbelot B., Arenas H., Cruz C. (2015). Lc3: A spatio-temporal and semantic model for knowledge discovery from geospatial datasets. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 35, p. 3–24.

Ireson N., Ciravegna F. (2010). Toponym resolution in social media. In *International semantic web conference*, p. 370–385.

Kim J., Vasardani M., Winter S. (2015). Harvesting large corpora for generating place graphs. In *International workshop on cognitive engineering for spatial information processes (cesip), in conjunction with cosit*, vol. 12.

Konvitz J. W. (1982). Redating and rethinking the cassini geodetic surveys of france, 1730-1750. *Cartographica*, vol. 28, p. 1–15.

Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., Mendes P. N. *et al.* (2015). Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, vol. 6, n° 2, p. 167–195.

Lesbegueries J. (2007). *Plate-forme pour l'indexation spatiale multi-niveaux d'un corpus territorialisé.* Thèse de doctorat non publiée, Université de Pau et des Pays de l'Adour.

Moncla L. (2015). *Automatic reconstruction of itineraries from descriptive texts.* Thèse de doctorat non publiée, Université de Pau et des Pays de l'Adour.

Moncla L., Renteria-Agualimpia W., Nogueras-Iso J., Gaio M. (2014). Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. In *Proceedings of the 22nd acm sigspatial international conference on advances in geographic information systems*, p. 183–192.

Moro A., Raganato A., Navigli R. (2014). Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, vol. 2, p. 231–244.

Nadeau D., Sekine S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, vol. 30, n° 1, p. 3–26.

OGC. (2011). *Ogc reference model*. Informative/Educational. The Open Geospatial Consortium (http://www.opengis.net/). (http://www.opengis.net/doc/orm/2.1)

Paris P.-H., Abadie N., Brando C. (2017). Linking spatial named entities to the web of data for geographical analysis of historical texts. *Journal of Map & Geography Libraries*, vol. 13, n° 1, p. 82–110.

Paulheim H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, vol. 8, n° 3, p. 489–508.

Peuquet D. J. (1994). It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of american Geographers*, vol. 84, n° 3, p. 441–461.

Puissant L. (1812). *Mémoire sur la projection de cassini, par l. puissant, pour servir de supplément à sa théorie des projections des cartes géographiques*. Paris, Vve Courcier.

Raymond C., Fayolle J. (2010). Reconnaissance robuste d'entités nommées sur de la parole transcrite automatiquement. In *Conférence traitement automatique des langues naturelles, taln'10*.

Rebele T., Suchanek F., Hoffart J., Biega J., Kuzey E., Weikum G. (2016). Yago: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *International semantic web conference*, p. 177–185.

Santos J., Anastácio I., Martins B. (2015). Using machine learning methods for disambiguating place references in textual documents. *GeoJournal*, vol. 80, n° 3, p. 375–392.

Speriosu M., Baldridge J. (2013). Text-driven toponym resolution using indirect supervision. In *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: Long papers)*, vol. 1, p. 1466–1476.

Stern R., Sagot B. (2010). Détection et résolution d'entités nommées dans des dépêches d'agence. In *Traitement automatique des langues naturelles: Taln 2010*.

Thury C.-F. Cassini de. (1783). *Description géométrique de la france*. impr. de J.-C. Desaint.

Tkachenko M., Simanovsky A. (2012). Named entity recognition: Exploring features. In *Konvens*, p. 118–127.

Vrandečić D., Krötzsch M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, vol. 57, n° 10, p. 78–85.

Weikum G., Hoffart J., Suchanek F. M. (2016). Ten years of knowledge harvesting: Lessons and challenges. *IEEE Data Eng. Bull.*, vol. 39, n° 3, p. 41–50.

Wu Y.-C., Fan T.-K., Lee Y.-S., Yen S.-J. (2006). Extracting named entities using support vector machines. In *International workshop on knowledge discovery in life science literature*, p. 91–103.

Zaveri A., Rula A., Maurino A., Pietrobon R., Lehmann J., Auer S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, vol. 7, n° 1, p. 63–93.

Zhao J., Jin P., Zhang Q., Wen R. (2014). Exploiting location information for web search. *Computers in Human Behavior*, vol. 30, p. 378–388.