

Création d'une base de connaissances sur les ports à partir des textes des instructions nautiques

Mots-clés: création et peuplement de bases de connaissances géoréférencées, intégration de données à références spatiales, extraction d'informations topographiques à partir de textes.

Contexte et objectifs

L'Institut national de l'information géographique et forestière (IGN) et le Service Hydrographique et Océanographique de la Marine (SHOM) sont les opérateurs publics respectivement en charge de l'information géographique et forestière et de l'information géographique maritime et littorale de référence. À ce titre, les deux établissements produisent des référentiels et des services destinés à répondre aux besoins d'informations géolocalisées, notamment au profit des politiques publiques comportant des enjeux d'analyse spatiale et de localisation. Ceci suppose de produire des référentiels qui soient à la fois les plus exhaustifs, détaillés, qualifiés, à jour et interopérables possibles.

Les référentiels de données géoréférencées sont de plus en plus utilisés pour permettre l'annotation spatiale de documents textuels et ainsi faciliter l'accès à leur contenu, voire son analyse spatiale. En revanche, peu de travaux se sont intéressés à l'extraction d'information géographique à partir de textes pour alimenter de tels référentiels. Pourtant, certains textes offrent des descriptions de l'espace très précises et détaillées et constituent parfois la seule source d'informations disponible.

Ce stage explorera les potentialités de l'extraction d'information à composante spatiale dans des textes allié aux standards du Web de données pour construire et peupler une base de connaissances sur les ports et mouillages décrivant leur localisation, leur configuration spatiale, leurs équipements, etc. à partir des *Instructions Nautiques* produites par le SHOM. L'objectif est de se doter de connaissances structurées pour répondre à des requêtes sur la localisation des ports et mouillages, leurs conditions d'accès et d'utilisation et à terme développer des services d'aide à la navigation à base de raisonnement.

Corpus de travail

Les *Instructions Nautiques* sont des documents textuels décrivant les amers et dangers pour la navigation côtière, les ports et mouillages, leurs chenaux d'accès, leurs équipements et les services proposés aux navigateurs, etc.

Les sections dédiées à la description des ports et mouillages suivent une structure relativement régulière. Elles débutent par un paragraphe décrivant la localisation générale du port, sa capacité d'accueil ainsi que son statut administratif. Puis viennent le plus souvent des explications sur les manoeuvres de chenalage pour y accéder, suivies d'une description des différentes zones de mouillage et installations portuaires. On y trouve enfin la liste des équipements - notamment à destination des plaisanciers - ainsi que les coordonnées des services utiles - bureau du port, mairie, délégation à la mer et aux affaires littorales, etc. L'extrait ci-dessous décrit le port de Saint-Cast.

01 **4.6.2.4. Saint-Cast**

07 L'Anse de Saint-Cast, ouverte au NE entre la Pointe de Saint-Cast et la Pointe de la Garde, est en partie ensablée et n'offre qu'un mauvais mouillage. Elle est bien protégée des vents, sauf de ceux du Nord à l'Est.

13 Le **Port de Saint-Cast** (48° 38' N — 2° 15' W) est adossé à la Pointe de Saint-Cast, au Sud. C'est un port départemental, de pêche et de plaisance, géré par la chambre de commerce et d'industrie des Côtes d'Armor. Il comprend un bassin à flot et une zone d'échouage.

19 Le bassin à flot, accessible à toute heure, est protégé à l'Est et au SE par une digue courbe dont le musoir porte un feu, et au SW par une ligne de pontons lourds marquée à son extrémité par une balise lumineuse.

25



4.6.2.4. — Saint-Cast, au NW (2007, avant transformation).

31 CHENALAGE. — L'approche se fait en venant du NE, en passant soit dans le NW, soit dans le SE du **plateau des Bourdinots**, marqué au NE par une bouée cardinale Est mouillée à 1 M au NE du port. De nuit, on fait route en se tenant dans l'un des deux secteurs blancs (206° – 217° ou 235° – 245°) du feu implanté à l'enracinement de la digue.

37 Un court chenal, à peu près parallèle à la jetée, conduit à la passe d'accès au bassin. Il est balisé par deux paires de bouées latérales lumineuses.

43 Le bassin à flot se compose de trois zones :

- au NE, une zone réservée à la pêche ; elle est équipée d'un ponton pouvant accueillir une vingtaine de bateaux, par profondeur de 2 m ;
- au centre, une zone dite « de grande plaisance », par profondeur de 2 m, destinée aux bateaux de 8 à 14 m de long ;
- au SW, une zone « de petite plaisance », par profondeur de 1 m, destinée aux bateaux de 5 à 8 m de long.

49 Le port d'échouage s'étend au SW du bassin. 19 mouillages sont réservés aux bateaux de pêche, au Sud de l'extrémité de la jetée.

Extrait des instructions nautiques pour le port de Saint-Cast (Source : SHOM)

Verrous scientifiques

Il s'agira d'extraire, typer, désambigüiser et structurer les informations sur les ports et les diverses entités spatiales qui les composent décrites par les textes (noms, types d'objets géographiques, localisations absolues et relatives, fonctions, ...) pour les intégrer dans une base de connaissances et vérifier la cohérence des informations extraites, inférer de nouveaux faits. Ceci suppose de proposer des solutions pour :

- Adapter les approches existantes d'extraction d'information à composante spatiale à partir de textes à des corpus techniques caractérisés par un vocabulaire très spécifique au domaine, mais

relativement peu structurés au sein des différentes sections. En particulier, de nombreuses entités spatiales mentionnées ne possèdent pas de nom ;

- Représenter, stocker et manipuler des informations à composante spatiale qualitative (références spatiales indirectes, positionnement relatif, etc.) et à différents niveaux de détail selon les standards du Web de données ;
- Désambiguïser les entités spatiales extraites. Ceci nécessite de prendre en compte :
 - ◆ des critères de liage dont la disponibilité pourra varier,
 - ◆ les éventuelles variations des entités spatiales d'une source à l'autre (variations de nom, de propriétés, de temporalité, de niveau de détail de descriptions, etc.).
- Détecter et corriger d'éventuelles incohérences spatiales ou temporelles dans les informations extraites, améliorer le typage des entités spatiales, inférer des relations spatio-temporelles entre entités géographiques, etc.

Renseignements pratiques

Une poursuite du sujet de stage en thèse de doctorat est possible (financement SHOM/IGN).

Profil recherché : Master 2 ou diplôme d'ingénieur en informatique : représentation de connaissances, Web sémantique, sciences de l'information géographique, extraction d'informations à partir de textes.

Compétences et connaissances:

- Un bon niveau en programmation est essentiel
- En raison de la nature du corpus de documents, la maîtrise du français est nécessaire.
- Extraction d'informations à partir de textes.
- Données géographiques vectorielles.
- Web de données, notamment RDFS et OWL.

Durée et période de stage : 5 mois, au cours du printemps et de l'été 2018.

Lieu du stage : Equipe [LaSTIG/Strudel](#), [Institut national de l'information géographique et forestière](#) (IGN), Saint-Mandé (métro 1, station Saint Mandé).

Indemnités de stage : Stage gratifié selon la législation française.

Modalités de candidature : Envoyer un CV, une lettre de motivation **ciblée sur le sujet** et les relevés de notes des 2 dernières années d'études par email, au format PDF et en un seul fichier à nathalie-f.abadie@ign.fr

Encadrement du stage :

- [Nathalie Abadie](#) (LaSTIG - COGIT/Strudel, IGN, nathalie-f.abadie@ign.fr)
- [Eric Kergosien](#) (Geriico - Université de Lille, eric.kergosien@univ-lille3.fr)
- [Eric Saux](#) (Irenav - Ecole Navale, eric.saux@ecole-navale.fr)