

Extraction de connaissances pour la description de l'environnement maritime côtier à partir de textes d'aide à la navigation

Léa Lamotte*, Nathalie Abadie*, Éric Saux**, Éric Kergosien***

* Univ. Paris-Est, LASTIG STRUDEL, IGN, ENSG, F-94160 Saint-Mandé
lea.lamotte@ign.fr, nathalie-f.abadie@ign.fr

<http://recherche.ign.fr>

**IRENav, Lanvéoc-Poulmic, CC 600, F-29240 Brest Cedex 9
eric.saux@ecole-navale.fr

<http://www.ecole-navale.fr>

***GERiico, Univ. de Lille, F-59000 Villeneuve d'Ascq
eric.kergosien@univ-lille.fr

<https://geriico-recherche.univ-lille3.fr>

Résumé. 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. Le travail présenté dans cet article explore les potentialités de l'extraction d'information spatiale indirecte (noms de lieux, relations spatiales, etc.) dans les textes des Instructions Nautiques produites par le Service Hydrographique et Océanographique de la Marine (SHOM). La méthode proposée combine une approche lexicale et une approche à base de patrons linguistiques puis est comparée aux principales approches d'extraction d'information géographique en français.

1 Introduction

Les Instructions Nautiques (IN) sont des documents textuels publiés par le Service Hydrographique et Océanographique de la Marine (SHOM), décrivant les amers et dangers pour la navigation côtière, les courants, les ports et mouillages, leurs chenaux d'accès, leurs équipements et les services proposés aux navigateurs, etc. L'extrait suivant présente une description d'un paysage typique des IN : "*[...]La Pointe de Pléneuf est débordée vers le NW par l'îlot Le Verdelet, puis jusqu'à 1,5 M au large par le Plateau des Jaunes, découvrant ; celui-ci, ainsi que les hauts-fonds rocheux qui s'étendent dans l'Ouest du Verdelet, sont couverts par le secteur rouge (146°-196°) du feu de la tourelle La Petite Muette située à l'entrée de Dahouët.[...]*".

Dans cet article, nous proposons une approche qui vise à extraire des Entités Spatiales Nommées (ESN) ou non et leurs relations spatiales à partir des IN en vue de les représenter au sein d'une ontologie. L'intérêt de ce modèle de représentation de connaissances est double.