

Vocabulary Control in Nautical Information Resources

Edgardo A. Stubbs

Instituto de Investigaciones en Humanidades y Ciencias
Sociales (IdIHCS), Universidad Nacional de La Plata

Abstract

Nautical charts are an essential information resource for safe navigation. However, they are not only a useful resource for navigators; According to the International Hydrological Organization (IHO), they essentially fulfill two functions: 1) Maritime navigation, since most hydrographic services have the obligation to provide coverage of the nautical charts of their national waters, in all coastal waters, including major ports and smaller marinas of purely local interest. and 2) as a source of information, since national nautical charts present a detailed configuration of the seabed. Information on the shape of the seabed is required by a diversity of users in addition to navigators; for example, engineers interested in onshore construction, dredging contractors, oceanographers, defense agencies, coastal zone managers, etc. Traditionally, there are three essential elements that play an important role in information retrieval: title, author, and subject access point. Among the latter, one can distinguish indexing by natural language and by controlled vocabularies. The thematic access points, makes it easier for the user to search and retrieve all types of resources that satisfy their information needs. Traditionally in the processing of nautical charts the natural language is used predominantly, motivated by a lack of availability of a controlled vocabulary specific to the area in Spanish. The objective of this work is aimed at establishing the criteria for the construction of a controlled vocabulary in Spanish in the field of nautical charts.

Keywords: nautical charts, controlled vocabulary, subject cataloguing, information needs, information retrieval, thesauri construction

Introduction

Nautical charts are a sort of cartographic information resources, which involve a graphical representation that provides highly significant information to ensure safe navigation. It indicates the depths of the sea, the nature of the seabed and the detailed maritime configuration. However, it is not only users of nautical charts who need to orient themselves to make a journey through navigable waters. The information provided by nautical charts exceeds the purpose for which they were originally

created. According to the Hydrographic and Oceanographic Service of the Chilean Navy (2003), a nautical chart is a graphic representation of the navigable waters and the adjacent coast drawn by means of an adequate projection system according to its purpose. It indicates the depth of the sea, the nature of the seabed and the detailed maritime configuration. It must also show the natural and artificial objects that exist fixed on land, the dangers to navigation, location of lights and other aids that are visible from the sea that can serve as a reference to the navigator to obtain their situation and draw directions.

Nautical charts in their evolution have been gaining importance over time not only as a support for navigation and maritime transport. They have also acquired importance for the determination of maritime safety processes in coastal procedures and in the prevention and monitoring of crisis situations created by ship accidents, in many cases with polluting or dangerous loads.

Nautical charts are also a tool to carry out environmental studies, studies related to natural resources, urban planning, tourism, etc.

In this sense, according to the Naval hydrography manual, the use of nautical charts is related to the following activities, namely

Maritime transport

More than 80% of the international trade in the world is transported by sea. Maritime trade is a basic element for the economy of a nation. Many areas and ports in the world do not have accurate or adequate coverage of their nautical charts. Modern nautical charts are necessary for safe navigation through the waters of a country and nearby coasts and for entry into its ports. The lack of adequate nautical charts prevents the development of maritime trade in the waters and ports of the nations in question. The shipping industry needs efficiency and safety. Poorly mapped areas and a lack of information can make trips longer than necessary, and can even prevent optimal loading of ships, thus increasing costs. The savings in time and money that result from the use of shorter and deeper routes and the use of larger ships or deep-draft freighters can represent important economic income for the industry and national commerce. It is also very important to note that Chapter V of the SOLAS (Save Our Life at Sea) convention considers a ship worthless if it does not carry the updated charts necessary for the intended voyage.

Coastal zone administration

Proper management of the coastal zone includes points such as the construction of new ports, and the maintenance and development of existing ones; dredging operations for the maintenance of charted depths and for the establishment, monitoring and improvement of channels; control of coastal erosion; land reclamation from the sea; for the establishment and monitoring of the guidelines that regulate the discharge of industrial waste; extraction of mineral deposits; aquaculture

activities; transportation and public works projects including construction of near-shore infrastructure.

The first essential data for the aforementioned projects are obtained from the coastal surveys carried out by the national Hydrographic services, the information of which is included in the corresponding nautical charts. Users of hydrographic information go beyond the traditional user group, mariners, including government agencies, coastal authorities, engineers and scientists.

Exploration and exploitation of maritime Resources

Although the information contained in the nautical charts, prepared by the national hydrographic offices, have been designed mainly to support the safety of navigation, such information also constitutes a considerable economic value for the administration and exploitation of the natural resources of the sea. For example the identification of minerals in marine sedimentary areas. The fishing industry also needs the information contained in the nautical charts, not only for the safety in the navigation of its vessels but also for the identification of the fishing areas.

The trend in modern fisheries science is toward habitat management, bathymetry, and other oceanographic data that will provide input of important data for proper species management and development.

Environmental protection and management

An essential factor for the protection of the environment is safe and accurate navigation. Waste and oil spills are a major destructive factor for pollution, the economic consequences are more devastating than imagined, in some cases, they have been estimated at 3,000 million dollars for a single incident. The value of shipping services for the protection of the marine environment has been recognized internationally. In this regard, it should be noted that Chapter XII of Agenda 21 of the United Nations Conference on Environment and Development (UNCED), held in 1992, recognized that "Nautical mapping is of vital importance for the navigation safety".

Methodology

The research methodology followed to address the issue raised has been based on a qualitative research process evaluating previous works of the author that allowed to establish the starting point of the present work

Subject cataloguing and information retrieval

Considering the cataloguing principles of 2009 and its modification of 2016 (IFLA, 2016), the catalog should be an efficient and effective instrument to find information resources that allow the user to identify a particular resource, a set of resources that belong to the same work, expression, manifestation; those associated with the person in charge of the information resource (s) or all the resources of a given subject, the latter subject of interest in this writing. From this possibility, the user will be able

to identify, select and retrieve the resource and / or related resources that he has been able to identify as a previous step to accessing the primary information. It is vitally important that the catalog, as far as the thematic description is concerned, enables the user to move (navigate) linking concepts, objects, events and places, constituting an element of selection, identification and information on the potentialities that the collection can provide. to satisfy your needs.

The cataloging principles establish that among the essential access points, controlled subject terms must be included as one of the necessary elements to enhance the identification and selection possibilities that the user has before going in search of the primary information resource.

A thesaurus according to ISO 2788-1986 defines a thesaurus as: "a controlled and dynamic vocabulary, composed of terms that have semantic and generic relationships between them and that is applied to a particular domain of knowledge". For his part, van Slype (1991) defines a thesaurus as: "a structured list of concepts intended to uniquely represent the content of documents and queries within a given document system and to help the user in the indexing of documents and queries."

Most authors agree that the first function of a thesaurus or any controlled language is to support the information retrieval process by constituting a guide that allows choosing terms for indexing and retrieving information.

According to ISO 25964-1 standard, The traditional aim of a thesaurus is to guide the indexer and the searcher to choose the same term for the same concept [...] a thesaurus should first list all the concepts that might be useful for retrieval purposes in a given domain. The concepts are represented by terms, and for each concept, one of the possible representations is selected as the preferred term [...] Secondly, a thesaurus should present the preferred terms in such a way that people will easily identify the one(s) they need. This is achieved by establishing relationships between terms — and/or between concepts — and using the relationships to present the terms in a structured display

In the information retrieval process prior to the development of thesauri, the use of character strings was common. With the advent of the thesaurus, the choice of preferred terms improved the information retrieval process

On the other hand, according to ISKO in the language of normal discourse one concept can be expressed in many different ways, and conversely one term can have many different meanings. To achieve the aim of always choosing the same term for the same concept an artificial indexing language has to be established, in which synonyms are controlled, homographs are disentangled, and each preferred term is allowed only one meaning (although some may have very broad meanings). The thesaurus conveys that artificial language. Thus the essential core of a thesaurus is a collection of concepts represented by terms and interlinked by relationships, of which the three main types are equivalence (between terms), hierarchical (between concepts) and

associative (also between concepts). By long established convention, the tags USE and UF (Use For) precede preferred and non-preferred terms respectively, and the equally characteristic tags BT, NT and RT indicate broader, narrower and associatively related concepts respectively.

The thesaurus can also be viewed as a complex network of interrelated concepts, in which each concept is labeled by one or more terms, in one or more languages, depending on whether it is a monolingual or multilingual thesaurus. These are some of the main characteristics

The semantic scope of a concept is indicated partly by the totality of terms labelling it, partly by the hierarchical relationships linking it to broader and/or narrower concepts, and where this is not enough, by a scope note and sometimes term definitions

Admissible hierarchical relationships are of three types: generic, instantial or partitive (subject to some restrictions on the eligible types of partitive link). It is optionally possible to distinguish these types, using the tags BTG/NTG, BTI/NTI, BTP/NTP respectively

Admissible associative relationships apply to non-hierarchical situations wherever two concepts are so associated that an indexer or a searcher should consider using one of them as well as, or instead of, the other

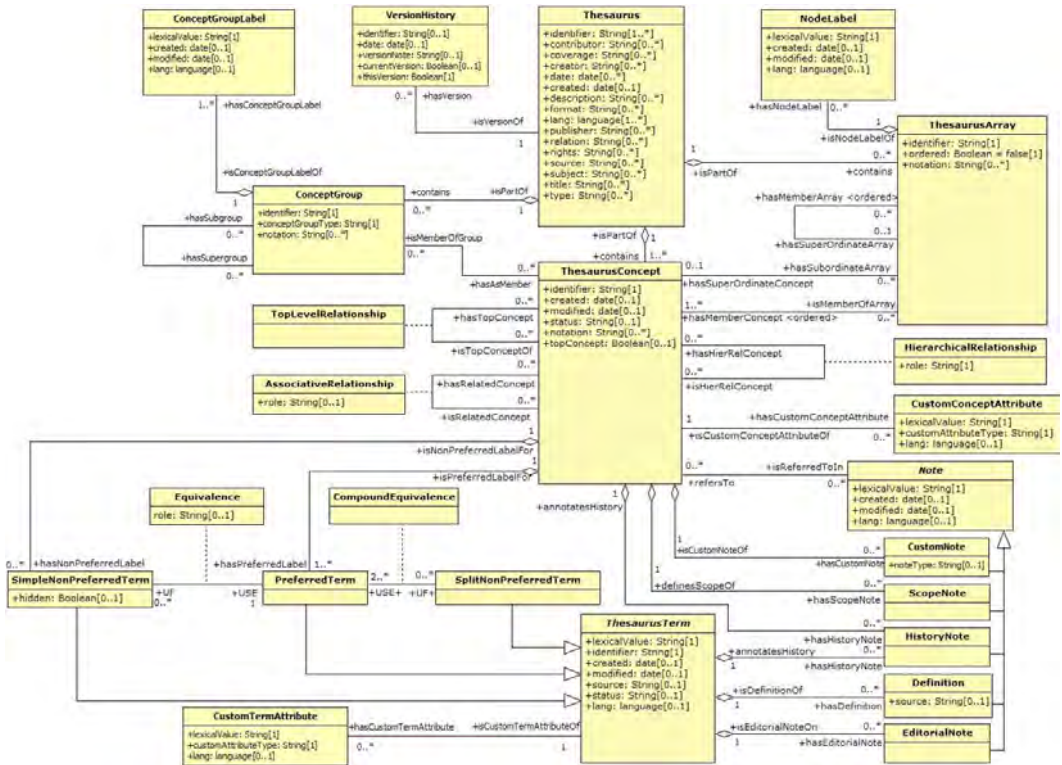
Concepts may be presented and ordered in arrays with node labels, following the principles of facet analysis

Concepts may also be grouped in loose structures to suit particular domains or applications

Concepts not explicitly included in the thesaurus may be represented by combinations of preferred terms (in situations known as "compound equivalence")

It is also possible to add metadata to terms, to concepts, to relationships and to the thesaurus as a whole, such as dates of introduction or change, version history, housekeeping data, copyright information, etc.

According to the ISO 25964-1 international standard ISO 25964-1 a thesaurus can be defined as a controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms.



A data model provided by ISO 25964-1 is shown

Subject Cataloguing and nautical charts

Subject description, through the thematic access points, makes it easier for the user to search and retrieve all kinds of documents that satisfy their information needs. According to (Stubbs et al., 2016) in the processing of nautical charts, the natural language is used predominantly, motivated by a lack of availability of a controlled vocabulary in Spanish specific to the area. The closest are vocabularies of general cartography that do not reach the specificity required for the description of the content of this class of resources. The general controlled vocabularies used in various information units and Libraries to process nautical charts, such as those with a certain degree of specificity, present important limitations for the description of content, making it necessary to include summaries or notes in bibliographic records. In four document languages used for the processing of nautical charts, it was observed that no selected vocabulary could represent at least 50% of the keywords selected for the content description in this class of information resources. (Stubbs, et. Al.; 2019)

It is important to note that many information agencies such as libraries and documentation centers have nautical charts given the interest and needs of the users that they must satisfy.

The control of vocabulary at the time of subject cataloging is essential for the retrieval of information in aspects such as the identification and link between related resources. Control of vocabulary in these sort of resource is related to general aspects of description, issues related to topography, underwater hazards, hydrographic data, and geographic location.

Given the particular characteristic of the cartographic resource and especially the nautical charts, one of the important aspects to take into account is the place or geographic location as terms to control for an optimization of the information retrieval.

Although in many cases the geographical location with which the user would go to look for the information is found or is part of the title, this is not always the case. And even in this case, the record would cease to fulfill one of the essential functions of the catalog, which is the link with other related documents, a situation that would not be exposed when describing only the document with the natural language, omitting the relationships and links that a cartographic resource may have with the rest. On the other hand, other nautical charts that are not linked to a specific locality or population are named by the geographical coordinates and since a geographical coordinate as a keyword is no longer very friendly. From this situation it is necessary to have a controlled vocabulary that in addition to univocally identify the terms and concepts that the resource identifies thematically (type of nautical chart, navigational warnings, natural or artificial underwater risks, aspect of the topography, etc.) It is important that the controlled vocabulary incorporates the standard terms of the location and its hierarchical relationships. This last aspect is what would allow the links to be established within the catalog between the different records that have some kind of hierarchical or associative relationship linked to geographic location.

Discussion

The lack of a controlled vocabulary in Spanish for the processing of nautical charts deprives the cataloger of a tool to incorporate one more element in information retrieval. Some studies suggest that the majority of searches were subject searches (Gross and Taylor, 2005). Gross and Taylor found that more than a third of records retrieved by keyword searches would be lost without subject headings. Which would indicate that not only is access by subject another element but a very important one that numerous studies, in various disciplines, have found that a quarter to a third of records returned in a keyword search would be lost without controlled vocabulary. So the lack of some kind of controlled vocabulary to retrieve information in relation to nautical charts directly affects the satisfaction of the user's needs in this area of knowledge.

Most of the studies carried out on information retrieval by subject have concluded that the greatest success in retrieving information is obtained with hybrid systems (systems that combine natural language with vocabulary control). According to

Lancaster (1991), systems that do not control vocabulary present problems when performing conceptual searches, so a combination of natural language and controlled language improves information retrieval.

Conclusion

Based on the foregoing, the processing of the information contained in nautical information resources (nautical charts) in Spanish presents weaknesses related to the lack of languages that allow controlling the vocabulary used when indexing an information resource.

As explained above, an attempt has been made to correct this deficiency from the use of natural language, which presents drawbacks related to the specificity of the terms with which the resource is described.

On the other hand, the lack of a controlled vocabulary makes the description present polysemics, ambiguities and inaccuracy in the results of an information search and at the same time it deprives one of the most important functionalities that controlled languages have is the link that allows establishing with similar resources by establishing hierarchical, associative and equivalence relationships between the terms

Finally, the development of a controlled vocabulary for the processing of nautical charts makes it possible to comply with one of the cataloging principles established in 2009, which implies the possibility that the user can navigate through the catalog.

The development of a controlled vocabulary in Spanish for the description of nautical charts constitutes an essential tool in improving information retrieval.

References

- [1] International Standart Organization (2011) ISO 25964. Thesauri and interoperability with other vocabularies. Geneva: ISO.
- [2] ISKO (2020) Thesaurus: for information retrieval. <https://www.isko.org/cyclo/thesaurus> [23-8-21]
- [3] SHOA. (2015). Catálogo de cartas y publicaciones náuticas. Valparaíso: SHOA. http://www.shoa.cl/s3/datos/catalogo_publicaciones/catalogo_03.pdf
- [4] SOLAS. (2015). Convenio para la Seguridad Humana en el Mar. http://www.inocar.mil.ec/web/images/lotaip/2015/literal_a/base_legal/A_Convenio_internacional_solas_1974.pdf [10-8-21]
- [5] Stubbs EA, Carut C, Gamba V, Mendes PV, Gutierrez C, Medina MC (2019) Descripción de contenido en el procesamiento de las cartas náuticas. Rev. Interam. Bibliotec (Colombia) 42(1) 37-44
- [6] United Nations (1992) Sustainable development. United Nations Conference on environment and development (Rio de Janeiro: June 3-14, 1992)

- [7] Van-Slipe, G . (1991). Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales. Madrid: German Sánchez Rupérez.
- [8] Tina Gross, Arlene G. Taylor & Daniel N. Joudrey (2015) Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching, Cataloging & Classification Quarterly, 53:1
- [9] Tina Gross and Arlene G. Taylor, "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results,"College & Research Libraries 66, 3 (2005): 212-30
- [10] Lancaster FW (1991) Indexing and abstracting in theory and practice. London:LA