

17 The population of a Maritime Heritage Thesaurus based on a Greek project documentation case

Eleni Georgaki, Sofia Stamou

ABSTRACT

This paper is related to the population of a thesaurus in the Maritime Heritage (MH) field. The suggested controlled vocabulary has been created to improve methods for archiving oral sources, written accounts, imagery, 3D archeology, and other multimedia objects related to people and periods from ancient times to recent history. It has been created to cover the indexing needs for resources from Greek Maritime GLAM institutions hosted in a MH platform of the research infrastructure EN.I.R.I.S.S.T. Its terms are 'derived from' the collections which include objects/heirlooms, personal items, scrimshaws, archaeological items, shipping companies' archives, correspondence, periodicals, photographs from coaling stations and lighthouses.

Although the construction of this vocabulary initially took into consideration specific data and their documentation, the goal is to go beyond EN.I.R.I.S.S.T., to assist in future MH documentation projects leaving an imprint in the improvement of MH data communication in general, and contributing to the promotion of MH as a separate Cultural Heritage (CH) branch. This is why, currently, the population of the first operational draft of the thesaurus is being attempted. Approximately 4,000 terms from maritime legislation documents, extracted automatically from an initial corpus of 80,000 words of the Government Gazette of the years between 1975-1999 will be incorporated in the MHThesaurus as a further development step.

Η συμπλήρωση ενός Θησαυρού Ναυτικής/Ναυτιλιακής Κληρονομιάς με βάση μια ελληνική περίπτωση έργου τεκμηρίωσης

Ελένη Γεωργάκη, Σοφία Στάμου

ΠΕΡΙΛΗΨΗ

Αυτό το άρθρο σχετίζεται με το γέμισμα ενός θησαυρού στον τομέα της Ναυτικής/Ναυτιλιακής Κληρονομιάς. Το προτεινόμενο ελεγχόμενο λεξιλόγιο έχει δημιουργηθεί για τη βελτίωση μεθόδων για την αρχειοθέτηση προφορικών πηγών, γραπτών λογαριασμών, εικόνων, τρισδιάστατης αρχαιολογίας και άλλων αντικειμένων πολυμέσων που σχετίζονται με ανθρώπους και περιόδους από την αρχαιότητα έως την πρόσφατη ιστορία. Έχει δημιουργηθεί για να καλύψει τις ανάγκες ευρετηρίασης για πόρους από ιδρύματα Ναυτικής/Ναυτιλιακής Κληρονομιάς (NNK), όπως Βιβλιοθήκες, Αρχεία και Μουσεία, που φιλοξενούνται σε πλατφόρμα της ίδιας θεματικής της ερευνητικής υποδομής EN.I.R.I.S.S.T. Οι όροι του «προέρχονται από» τις συλλογές που περιλαμβάνουν αντικείμενα/κειμήλια, προσωπικά αντικείμενα, εγχάρακτα αντικείμενα, αρχαιολογικά αντικείμενα, αρχεία ναυτιλιακών εταιρειών, αλληλογραφία, περιοδικά, φωτογραφίες από ανθρακευτικούς σταθμούς και φάρους.

Η έρευνα που παρουσιάζεται είναι η δημιουργία ενός θησαυρού που περιγράφει επίκαιρα θέματα που απεικονίζονται οπτικά σε έγγραφα, φωτογραφίες και αντικείμενα NNK κληρονομιάς, αλλά έχει προβλέψει επίσης την ολοκλήρωση άλλων πεδίων του Dublin Core στη διαδικασία καταλογογράφησης, όπως η «Κάλυψη» και το «Μορφότυπο». Μπορεί να θεωρηθεί ως «κέλυφος» και σίγουρα όχι ολοκληρωμένο έργο αφού η προσθήκη νέων ψηφιοποιημένων και διασυνδεδεμένων στοιχείων θα οδηγήσει σε επακόλουθη επέκταση του θησαυρού και αυτή η διαδικασία εκτιμάται ότι θα είναι συνεχής.

Για την κατασκευή ενός αυστηρά ελεγχόμενου λεξιλογίου ακολουθήθηκε ως διαδικασία η χρήση μιας βελτιωμένης πολυεπίπεδης ιεραρχικής διάταξης και η τοποθέτηση συγκεκριμένων εννοιών εντός της ιεραρχικής δομής της ορολογίας που λειτουργούν ως βασικά σημεία πρόσβασης στις NNK συλλογές. Η βάση ανάπτυξης αυτού του έργου είναι η τεκμηρίωση του ψηφιοποιημένου πολιτιστικού και αρχαιολογικού υλικού που παρέχεται από ελληνικούς ναυτιλιακούς φορείς και τα δεδομένα τους εμπίπτουν στις προαναφερθείσες κατηγορίες δεδομένων της πλατφόρμας.

Η τυπολογία των αντικειμένων περιλαμβάνει αντικείμενα/κειμήλια που περιλαμβάνουν προσωπικά αντικείμενα, συλλογή αρχαιολογικών αντικειμένων, αρχεία ναυτιλιακών εταιρειών, προσωπική και επίσημη αλληλογραφία, περιοδικά, φωτογραφικό υλικό από ανθρακευτικούς σταθμούς και φάρους.

Η πλατφόρμα NNK, με τη συντομογραφία/ακρωνύμιο ΝΑΥΚΛΗΡΟΣ, αποτελεί πύλη πρόσβασης σε ψηφιοποιημένα τεκμήρια στο πλαίσιο της δημιουργίας της ηλεκτρονικής Εθνικής Ερευνητικής Υποδομής EN.I.R.I.S.S.T. με στόχο την προβολή τους, τη διασύνδεση με πηγές αναφοράς και την προσέλευση περαιτέρω διασυνδέσεων με άλλες ηλεκτρονικές πλατφόρμες παρόμοιου περιεχομένου τόσο στην Ελλάδα όσο και στο εξωτερικό. Η EN.I.R.I.S.S.T., στην οποία εντάσσεται η πλατφόρμα είναι μια μοναδική και πρωτοποριακή Εθνική Ερευνητική Υποδομή που υποστηρίζει και προάγει την έρευνα και έχει όραμα να αποτελέσει ένα κέντρο αριστείας στους τομείς της Ναυτιλίας, των Μεταφορών και της Εφοδιαστικής Αλυσίδας στην Ελλάδα. Η δημιουργία περισσότερων υπηρεσιών και η διασύνδεση με άλλα κύρια εθνικά και ευρωπαϊκά δίκτυα πολιτιστικής και ναυτιλιακής κληρονομιάς γίνεται στο πλαίσιο του διάδοχου έργου EN.I.R.I.S.S.T. Plus.

Αν και η κατασκευή αυτού του λεξιλογίου αρχικά έλαβε υπόψη συγκεκριμένα δεδομένα και την τεκμηρίωσή τους, ο στόχος είναι να υπερβεί το EN.I.R.I.S.S.T., να βοηθήσει σε μελλοντικά έργα τεκμηρίωσης Ναυτικής κληρονομιάς αφήνοντας ένα αποτύπωμα στη βελτίωση της επικοινωνίας τους γενικά και συμβάλλοντας για την ανάδειξη της NNK ως ξεχωριστού κλάδου Πολιτιστικής Κληρονομιάς.. Πέρα από το EN.I.R.I.S.S.T., μακροπρόθεσμα μπορεί επίσης να βοηθήσει έργα τεκμηρίωσης της NNK ή των επαγγελματιών των ναυτιλιακών βιβλιοθηκών, αρχείων και μουσείων στην καθημερινή τους εργασία ή απλώς να αποτελέσει αφηρητά για ερευνητές που ενδιαφέρονται να οργανώσουν και να αρχειοθετήσουν πολιτιστικά δεδομένα σχετικά με την NNK. Η φιλοδοξία σίγουρα δεν είναι όλα τα έργα, ψηφιακά ή άλλα, να αξιοποιηθούν αυτό το λεξιλόγιο, αλλά έχουμε την ελπίδα να ενθαρρύνουμε τον προβληματισμό σχετικά με τον τρόπο περιγραφής και οργάνωσης των δεδομένων.

Γ' αυτό, επί του παρόντος, επιχειρείται το γέμισμα της πρώτης εκδοχής του θησαυρού. Περίπου 4.000 όροι από έγγραφα της ναυτικής νομοθεσίας, που εξήχθησαν αυτόματα από ένα αρχικό σώμα 80.000

λέξεων της Εφημερίδας της Κυβερνήσεως των ετών 1975-1999 θα ενσωματωθούν στο MHThesaurus ως περαιτέρω αναπτυξιακό βήμα.

0 Introduction

The current introduction is related to the creation of a thesaurus in the Maritime Heritage (MH) field which is certainly sought after and innovative, as there is no precedent. The suggested controlled vocabulary could definitely improve methods for archiving oral sources, written accounts, imagery, 3D archeology, and other multimedia objects related to people and periods from ancient times to recent history. It has been created to cover the particular indexing needs for museum and archive resources from Greek Maritime GLAM institutions hosted in a MH platform of the research infrastructure EN.I.R.I.S.S.T. It is comprised of terms 'derived from' the collections which include scans of models and designs that are related to emblematic categories / types of ships-boats historically, digital files of sailing boats images related to categories of sailing boats, scans of navigation instruments and related items, other museum objects of maritime museums and organisations and material related to the history of shipping (testimonies, etc.).

The research is primarily a practical design of a thesaurus which describes topical subjects visually represented in MH documents, photographs, and artifacts, but also has predicted the completion of other Dublin Core fields in the cataloguing process, such as the 'Coverage' and the 'Format' fields. It can be regarded as a 'shell' and is definitely not a complete work since the addition of new digitized and interconnected items will lead to a subsequent expansion of the thesaurus and this process is estimated to be continuous.

For the construction of a strictly controlled vocabulary the use of a streamlined multi-tiered hierarchical arrangement and the placement of specific concepts within the hierarchical structure of terminology that function as key access points to MH collections was followed as a procedure. The basis of development of this project is the documentation of the digitized and digitalized cultural and archival material provided by Greek maritime institutions-providers and their data fall into the previously mentioned data categories of the platform.

The typology of the objects includes objects/heirlooms that incorporates personal items, collection of scrimshaws, collection of archaeological items, archives of shipping companies, personal and official correspondence, periodicals, photographic material from coaling stations and lighthouses.

1 State of the art

At present, the e-documentation of digital cultural heritage objects is inadequate to provide knowledge and data needed by multiple possible audiences, and as a result, professionals, experts and the general public accessing major data aggregators such as Europeana, the EU digital cultural heritage library, are not able to discover this information (Ioannides & Davies, 2019, p. 76-7). The enhancement of facilities of institutions in the GLAM sector for exposing, searching and sharing the knowledge implicit in their collections is an ongoing research interest pictured in the academic bibliography. During the recent decades semantic web technology has been leveraged extensively to enable retrieval, navigation, browsing and effective annotation of the museum collections, as well as an effective platform for interoperability with other systems and a potential base for reasoning and data mining. Ontologies function as a basis for visualising the inherent knowledge in the collections and the development of what is known as the semantic layer above the raw data. This semantic layer provides through associations with the media objects the explicit conceptualisations of the domain and together they facilitate semantic navigation, browsing and retrieval as new ways for exploring collections (Goodall et al, 2004). Semantic interoperability is a prerequisite for knowledge discovery, inference, and data federation between information systems, particularly when data sources from different providers need to be combined (Fafalios et al., 2021).

Recent research states that digitization paces are still lower than needed and the proportion of the digital CH objects in the European institutions to which descriptive metadata have applied is around 50%. Given the heterogeneous contents and formats of data (3D/2D, textual, audio, video, multilingual) quality standards to ensure findable, accessible, interoperable, and reusable (FAIR) data is essential to become a priority so as to achieve the desired level of integration, enrichment, retrieval and reuse of the content (Ioannides & Davies, 2019, p. 77-9· Castelli et al, 2021). Through the way of standardization of their data and metadata CH institutions will reach not only interoperability but also connection to geo-spatial, bibliographic and archival data offering a more holistic documentation. Linked (Open) Data perform a critical role in this kind of digital transformation of collections, and to be applied they require, semantics-ontologies and mapping models, controlled vocabularies. To share LOD data their standardization covering core concepts such as people, places, events etc is demanded and succeeded when developing and promoting the corresponding frameworks. To make the most out of technology organizational upgrade of CH institutions is a prerequisite (ibid).

The existing previous literature is really rich regarding the thesauri and controlled vocabularies since they exist across the disciplines. Hedden (2008) delineates the framework of controlled vocabularies distinguishing them from taxonomies and thesauri pointing to specific requirements and indexing purposes to be covered despite the blurring of the names. Controlled vocabularies may be used as the 'umbrella' term explaining the way of the construction of these vocabularies since contributors/creators and the way of structure is set. But it can also refer to the list of words used to index. Actually, they are an organized arrangement of words and phrases used for a twofold purpose; to index content and/or to retrieve content through browsing or searching, typically including preferred and variant terms and having a defined scope or describing a specific domain, and their types are not mutually exclusive (Harpring, 2010, p. 12 & 18).

On the other hand, a thesaurus is defined as a kind of dictionary that contains synonyms or alternative expressions for each term, and possibly even antonyms but also provides information about each term and its relationships to other terms within the same thesaurus including broader, narrower and related terms. Its building is complied with national and international standards, such as ISO 2788, ISO 5964, ANSI/NISO Z39.19 (Hedden, 2008, p. 33). They may be monolingual or multilingual and generally used for browsing in post-coordinated information storage and retrieval systems (Harpring, 2010). Thesauri standards are designed to make the thesaurus work reliably as a guide for choosing the right term for the concept sought by clarifying their interrelationships with others (Clarke & Zeng, 2012). The success of a thesaurus is directly dependent on its ability to "copy" the way the human brain functions; a huge and complex web of interdependencies is stored in the human mind within which no word is "in isolation island". Each thought, word or image is associated with other related words and concepts through subtle relationships (Giannibas & Brindesi, 2006). The methodology used in the analysis of terminologies requires identification of the context or subject area under consideration, identification of the properties attributed to the objects within the subject field, identification of those properties that are abstracted and are understood as characteristics and then combining the characteristics to form one concept. Since the terms always deal with the special language of a particular field of knowledge, the concept must be treated not only as a unit of thought but also as a unit of knowledge and most of the times extensive research work is required. Its depth and width should be explored and for concept system modeling hierarchical, special, divisive and associative relationships must be used (ELOT 402, 2008).

Semantic manipulation and the need for machine-to-machine communication, a demand for interoperability to underpin web services, publishing, aggregation, and the exchange of thesaurus data via multiple media and formats, expansion of searches across networked repositories are pressing needs that thesauri and other types of controlled vocabularies serve (Clarke & Zeng, 2012· Fafalios et al., 2021).

In essence, maritime collections and archival documents in maritime history are related to: maritime engineering, naval architecture, trade and marine transportation, oceanography, transportation economics and management, water activity in general. The scope of these special maritime collections might include journals and proceedings of professional maritime organizations, genealogical records, collections of journals reports, daybooks and personal letters, company documents of shipping companies and cruise liners. In addition, maritime heritage collections contain museum collections including artifacts from ships and life at sea, such as ship models and nautical equipment or works of art as well as photographs and miscellaneous documents detailing the history of ships and maritime activities (Constantinou, 2013).

Maritime heritage is mainly connected to the Maritime museums collections as well as other institutions of maritime tradition and culture. In Greece, it is documented in the museums and libraries that have relevant documents such as the Maritime Libraries Network (Aegean University Library, Piraeus University Library, NTUA Library, Naval Cadet School Library, Aikaterini Laskaridis Library, Eugenides Foundation Library, Maritime Museum of Greece) (Maritime Libraries Network, n.d.) and in the collections of maritime museums, libraries and maritime tradition institutions. Archaeological objects, historical writings and manuscripts, folklore objects and relics, works of fine arts, cartographic materials, cultural material of intangible heritage, books, electronic material, maps, coat of arms, Doctoral Dissertations, naval designs, photos fall into the typology of Greek maritime heritage.

There are many challenges encountered when trying to create subject-oriented controlled vocabularies like subject headings lists or thematic thesauri. Boulay (2006, p. 17) reports *ambiguous terminology, lengthy list with swallow hierarchical division and the inability to catalogue properly* as well as budgetary constraints-noted in our case as well- which do not permit the luxury of ideal choices and create a lot of room for compromise. The researcher summarizes the essence of such attempt as very complicated since such a cataloguing and research tool has to be *painstakingly tailored to each area of subject specialization prior to its implementation* (Boulay, 2006, p. 29).

There have been previous attempts of building topical thesauri in many areas such as History, where Boulay (2006) described war photographs by designing a list of subject headings in her thesis. The importance of this research was special as photography as a type of evidence is found to a significant extent in various branches of Cultural Heritage. A similar research to (Georgaki, thesis in progress), in terms of subject, was conducted by Borsato (2021) who studied ancient Catalan texts of a practical nature and collected new technical lexicon in Catalan to compare it with the Venetian, entered in a local scientific database. Contributing to the enrichment of the lexical heritage with maritime and naval vocabulary the researcher created a glossary. Ontology-based visualizations of cultural heritage have been provided by many researchers (e.g. Damiano et al., 2014), and there are plenty of research leveraging automatic extraction of terms from corpora in various fields (e.g. Mouratidis et al., 2022; Ataeva et al., 2019) apart from the specialized controlled vocabularies carried out by institutional documentation centres, such as the International Bureau of Education (IBE), contributing the UNESCO : IBE Education Thesaurus.

2 Methodology

2.1 Method: The terms and their organization

Developing a MH controlled vocabulary requires a methodological approach that must be sensitive to the history of the artefacts, the shipping history and terminology, cultural heritage vocabularies and general vocabularies, as well as the pre-existing practices. The procedure involved two stages: first the monolingual thesaurus was produced, and then the translation of the reduced thesaurus was carried out, sometimes, taking into account terms and subject headings already existing in the vocabularies of reference.

To create the thesaurus, the rules and standards of thematic indexing, creation of monolingual and multilingual thesauri, as well as terminology standards were followed.

For the construction of a strictly controlled vocabulary the use of a streamlined multi-tiered hierarchical arrangement and the placement of specific concepts within the hierarchical structure of terminology that function as key access points to MH collections was followed as a procedure. The basis of development of this project is the documentation of the digitized and digitalized cultural and archival material provided by the maritime institutions-providers. The specifications of the corresponding ELOT and ISO standards were followed as well as the international developments in the matters of construction and management of treasures, but also the Greek peculiarities in the issues of language and terminology. For instance, the National Library of Greece subject headings, as well as definitions from the Centre for the

Greek Language in order to guide the cataloguer or the user in the 'Notes' section were adopted or seriously taken into consideration.

The vocabulary developed is subject-specific and based on pre-existing controlled vocabularies in the CH domain or general, which include terms of the particular field. Excerpts from lists of thematic headings which are considered to contribute were also studied and adopted, or adapted accordingly, in some cases, as well as incorporation of terminology reported in literature.
















Following this methodology, an initial set of top-level concepts was primarily designed based on other previously mentioned vocabularies but also creating a first operational draft. Thesauri cooperation and terminology integration are highly intended to ensure reliability of the endeavor and ensuring that previous cataloguing practice was integrated at least as far as the first experimental set of data is concerned, and when put in practice new entries and finer ramifications may apply.

The research is primarily a practical design of a list of subject headings which describes topical subjects visually represented in MH documents, photographs and artifacts, but also has predicted the completion of other Dublin Core fields in the cataloguing process, such as the Coverage and the Format.

This thesaurus is not merely restricted to the description of the subject matter, although this is predominant. This was not originally intended but it came up as a need because of the existence of unique in the MH field cultural objects, such as scrimshaws. As such, the list does not replace the role of other fields that need to be catalogued such as: type, format, coverage, historical period, creator, contributor, title, or language but it is carried out to be supplement preexisting vocabularies, or simply assist the process of documenting the previously mentioned fields. The controlled vocabulary is restricted as much as possible to visually discernable subjects, and avoids the use of concepts and subjects related to the context of the visual item/photograph. The suggested hierarchical structure minimizes ambiguous or synonymous concepts. Four (4) facets and eleven (10) descriptors/identifiers have been singled out as access points to MH collections and are placed in the first tier of the list to successfully narrow down the scope of searching. As shown in Figure 1, there are the following facets: Concepts, Material objects, Activities, Geopolitical units, and descriptors/identifiers: Economy, Science, Culture, Information and Communication, Item types, Countries and country groupings. The "Orphans" facet and hierarchy include terms

that have been linked to them descriptors with an association relationship but not yet organized hierarchically.

Figure 1: facet search results

Αποτελέσματα Αναζήτησης Θεμάτων		Κριτήρια Αναζήτησης Θεμάτων
Στατιστικά: βρέθηκαν 5 αποτελέσματα. Αποτελέσματα: 1 - 5 σελ.: 1 / 1		
Θέμα	Ιεραρχίες	Ενέργειες
Γεωπολιτικές ενότητες (Geopolitical units)	Χώρες και ομαδοποίηση χωρών	  
Δραστηριότητες (Activities)	Υπαίθριες δραστηριότητες	  
Έννοιες (Concepts)	Δίκαιο, Επιστήμη, Οικονομία, Πληροφορία και επικοινωνία, Πολιτισμός, Τύποι τεκμηρίων	  
ΟΡΦΑΝΟΙ ΟΡΟΙ	Ορφανοί όροι	  
Υλικά αντικείμενα (Material objects)	Προσωπικά αντικείμενα, Φορτίο	  

As regards the construction of the facets, the process of their building was inspired by the Art and Architecture Thesaurus Facet code and the Backbone Thesaurus, a meta-thesaurus, which proposes a common model of thesaurus building.

As mentioned, this thesaurus is designed to be used at the MH platform NAFKLIROS of the research infrastructure EN.I.R.I.S.S.T., but definitely not limited. NAFKLIROS (NAFKLIROS / ΝΑΥΚΛΗΡΟΣ = Greek for “Bosun”) is the acronym of a maritime and naval heritage visualization platform providing access to reference resources and emblematic digitized maritime heritage artifacts of Greece while inviting further connection with digital platforms of a similar orientation both in Greece and Europe/internationally. The thousands of years long Greek maritime and naval heritage have bequeathed a very large number of related artifacts to the country’s organized collections a small percentage of which only is available in digital form and accessible electronically. The platform provides a general categorization of these in nature and a fitting ontology for search and retrieval purposes (NAFKLIROS, 2023).

The thesaurus is comprised of subject headings derived from the collections of SHIPMARK which includes scans of models and designs that are related to emblematic categories / types of ships-boats historically, SHIPSAIL which includes data digital files of sailing boats images related to emblematic categories of sailing boats, SHIPEQUIP which includes scans of navigation instruments and related items, RELATED MUSEUM ITEMS, which concerns other museum objects of maritime museums and organizations and SHIPSTORIES with other material related to the history of shipping (testimonies, etc.).

For the construction of a strictly controlled vocabulary the use of a streamlined multi-tiered hierarchical arrangement and the placement of specific concepts within the hierarchical

structure of terminology that function as key access points to MH collections was followed as a procedure. The basis of development of this project is the documentation of the digitalized cultural and archival material provided by the maritime institutions-providers. These are objects/heirlooms that incorporate personal items (Hellenic Maritime Museum), collection of scrimshaws (Hellenic Maritime Museum, Laskaridis Foundation, the Maritime museum of Oinousses), collection of archaeological items (the Maritime museum of Oinousses), archives of shipping companies (the Piraeus Bank Group Cultural Foundation), personal and official correspondence, periodicals (the Naval History Service), photographic material from coaling stations and lighthouses (the Photography Archive of the municipality of Kea island of Cyclades).

Lots of challenges were encountered during the construction of the thesaurus; not established/authorized way of cataloguing from the institutions to compare or base the research on was a challenge encountered by leveraging other aids to the building of the vocabulary. Finally, the thesaurus developed is subject-specific, faceted, and based on the Getty Foundation's Art & Architecture Thesaurus - AAT and Wikipedia's DBpedia semantic thesaurus, the established NLG subject headings, Thesaurus of Graphic Arts (TGA-LOC). Excerpts from lists of thematic headings (<https://southampton.spydus.co.uk/>, <https://biblionet.gr/>, <http://mobius.mysticseaport.org/>) were studied during the decision making about the final terms; The current proposed version of the vocabulary was based to a large extent on the UNESCO thesaurus and its Greek adaptation through a bilingual vocabulary of thematic tags which is complementary to the UNESCO Thesaurus carried out by the NDC in Greece. The choice is not at all arbitrary, but it serves interoperability without sacrificing precision with regard to the particular thematic area. Also, research in specific trade unions sources was conducted in order to effectively grasp the occupational activities and reach specific terminology enriching related terms and describing holistically terms that should be described (e.g. Metal industry).

The thesaurus was implemented with the Web-TMS (Thesaurus Management System) software supported by the Cultural Informatics research group of the Institute of Computer Science – FORTH, University of Crete. This system works in a graphical internet environment and provides for the construction and management of bilingual multi-thematic thesauri, in this case with Greek as the dominant language and English as the reference language.

The thesaurus can be found at <https://demos.isl.ics.forth.gr/themas-el/Index?logout=true&lang=en> with username MHTanagnostis and MHTanagnosi as a

password by selecting MHThesaurus and the link to the platform is <https://nafkliros.enirisst.gr/en/museums> with the respective documentation.

2.2 Validation

The thesaurus currently comprises of 158 terms in Greek and 140 in English. The hierarchies gathering most of the terms are 'Economy', 'Item types', 'Science' and 'Culture' with 56, 31, 21 and 13 terms respectively, whereas the vast majority (137 terms) belong to the 'Concepts' facet.

Experts, cataloguers from institutions included in the EN.I.R.I.S.S.T. project have been asked to review one or more hierarchies of their choice and comment on their structure, content, or translation. A questionnaire is being completed and used as a measuring instrument; it is comprised of three sections. The first section examined the competence of the introduction, the second one reviewed compliance with the thesaurus building criteria as derived from the literature review, and especially described in Spiteri (2000) and Lacasta et al. (2016), and the third examined particular terms, hierarchies and facets of the first operational draft of the thesaurus testing structure, content, translation and related elements, as appearing when the user browses the thesaurus online, or the Maritime Heritage platform of the EN.I.R.I.S.S.T. infrastructure selecting to view the metadata of items included in its collections.

The process is still ongoing, however some results have been already gathered and analyzed. Regarding the first section of the questionnaire, the majority of the answers provided commented that the introduction/description of the thesaurus which serves as a guide for the user explains clearly and in detail the facet principles and processes used to construct the thesaurus. There were also suggestions to make the project understandable without many technical details, or articulated in the logic of layers that open up for anyone who wants to learn more.

As for the second section examining the faceted display of the thesaurus, the questions in the various versions of the questionnaire were many looking into every single criterion when building a thesaurus, such as the presence, or not, of plurals, uppercase letters, prepositional sets, duplicates etc. We refer to the most essential and overarching questions regarding the facets and terms, and their relationships as a sample; 'Broad facets are mutually exclusive', 'broad facets are subdivided into top-terms', 'top-terms are labelled clearly', 'the relationships between the terms stand out', 'the connection of broader and narrower terms is shown and the corresponding "circles" are created in the terms', 'semantic

correctness of broader and narrower terms is observed' were reported positively 'most of the times', or 'always' by the respondents.

As far as the third section is concerned, valuable feedback was given regarding unclassified terms, suggestions of terms missing, or agreement/disagreement to the broader/narrower terms relations, as well as synonyms, and positive feedback is also provided on sample documentation, with few suggestions for adjustments.

3 Conclusions

The population of the thesaurus with terms from maritime legislation documents is currently designed. Authors in (Mouratidis et al., 2022) and (Mouratidis et al., 2023) have contributed to the advancement of deep learning applications in maritime terminology classification by developing a deep learning technique for text mining and discovering knowledge from textual data. They focused on legal texts for extracting concepts from the broader domain of maritime heritage. The results of their research, approximately 4000 words, extracted from an initial corpus of 80.000 included in legal texts on maritime matters in the Government Gazette, between the years 1975-1999, will be incorporated in the MHThesaurus as a further development step.

References

- Ataeva, O.M., Serebryakov, V.A., & Sinelnikova, E.K. (2019). Thesaurus and Ontology Building for Semantic Library Based on Mathematical Encyclopedia. *Proceedings of the XXI International Conference DAMDID / RCDL'2019*, 36 (1), 178-87. <https://ceur-ws.org/Vol-2523/paper15.pdf>
- Borsato, E. (2021). *Il lessico della navigazione e delle maestranze nella Venezia del XIV-XV secolo*. Studio a partire da alcuni zibaldoni marittimi [10.25434/borsato-ester_phd2021].
- Boulay, M.D. (2006). *Description of war photographs: designing a list of subject headings*. Toronto Metropolitan University. Thesis. https://doi.org/10.32920/ryerson.1464456_0.v2
- Castelli, L., Felicetti, A., & Proietti, F. (2021). Heritage Science and Cultural Heritage: standards and tools for establishing cross-domain data interoperability. *International Journal on Digital Libraries* 22, 279–287. <https://doi.org/10.1007/s00799-019-00275-2>
- Clarke, S.G., & Zeng, M.L. (2012). Standard Spotlight: From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards Towards Interoperability and Data Modeling. *Information Standards Quarterly*, 24, 20-26.
- Constantinou, C. (2013). Digitizing rare maritime collections. Στο *Archiving Conference, Archiving 2013 Final Program and Proceedings*, pp. 93-97.
- Fafalios, P., Petrakis, K., Samaritakis, G., Doerr, K., Kritsotaki, A., Tzitzikas, Y., & Doerr, M. (2021). FAST CAT: Collaborative Data Entry and Curation for Semantic Interoperability in Digital Humanities. *J. Comput. Cult. Herit.* 14(4), Article 45 (December 2021), 20 pages. <https://doi.org/10.1145/3461460>

Georgaki, E. (χ.χ.). Digital visual documentation and communication of Greek maritime heritage [thesis in progress]. University of the Aegean.

Giannibas, D., & Brindesi, H. (2006). Thesauri and information retrieval systems in strong interaction. In *Panhellenic Conference of Academic Libraries. TEI of Athens*. <http://abekt.lib.ucy.ac.cy/synedria/14psab/14psab032.pdf>

Harpring P. (2010). *Introduction to controlled vocabularies: terminology for art architecture and other cultural works* (1st ed.). Getty Research Institute.

Ioannides, M., & Davies, R. (2019). Towards a Holistic Documentation and Wider Use of Digital Cultural Heritage. *Communications in Computer and Information Science*, 846, 76-88. DOI: 10.1007/978-3-030-14401-2_7

Lacasta, F., Falquet, G., Zarazaga-Soria, J., & Nogueras-Iso, J. (2016). An automatic method for reporting the quality of thesauri. *Data & Knowledge Engineering*, 104, 1–14. <http://dx.doi.org/10.1016/j.datak.2016.05.002>

Mouratidis, D., Kermanidis, K., & Kanavos, A. (2023). *Comparative Study of Recurrent and Dense Neural Networks for Classifying Maritime Terms*. 978-1-6654-0032-9/21/\$31.00 © 2023 IEEE

Mouratidis, D., Mathe, E., Voutos, Y. Stamou, S., Kermanidis, K., Mylonas, P., & Kanavos, A. (2022). Domain-Specific Term Extraction: A Case Study on Greek Maritime Legal Texts. Στο *Proceedings of the 12th Hellenic Conference on Artificial Intelligence (SETN '22)*. Association for Computing Machinery, New York, NY, USA, Article 10, 1–6. <https://doi.org/10.1145/3549737.3549751>

NAFKLIROS. (2023). <https://nafkliros.enirisst.gr/en>

Spiteri, L.F.. (2000). The Essential Elements of Faceted Thesauri. *Cataloging & Classification Quarterly*, 28(4), 31-52. DOI:10.1300/J104v28n04_05.

THEMAS -Thesaurus Management System. (2020). Information Systems Laboratory (ICS-FORTH). <https://www.ics.forth.gr/isl/THEMAS/>

Ελληνικός Οργανισμός Τυποποίησης. (2010). *Ορολογική εργασία – Αρχές και μέθοδοι*. (ΕΛΟΤ 402:2010, 2^η έκδ.)

Appendix

English-Greek Glossary

activities	δραστηριότητες
broader term	πλατύτερος όρος
concepts	έννοιες
countries and country groupings	χώρες και ομαδοποίηση χωρών
culture	πολιτισμός
cultural heritage (CH)	πολιτιστική κληρονομιά
economy	οικονομία
facet	θέμα

geopolitical units	γεωπολιτικές ενότητες
galleries, libraries, archives, and museums (GLAM)	πινακοθήκες, βιβλιοθήκες, αρχεία, και μουσεία
information and communication	πληροφορία και επικοινωνία
Intelligent Research Infrastructure for Shipping, Supply Chain, Transport and Logistics (EN.I.R.I.S.S.T.)	Ευφυής Ερευνητική Υποδομή στη Ναυτιλία, την Εφοδιαστική Αλυσίδα, και τις Μεταφορές
item types	τύποι τεκμηρίων
linked (open) data (LOD)	ανοιχτά διασυνδεδεμένα δεδομένα
Maritime Libraries Network (MarLiNet)	Δίκτυο Ναυτιλιακών Βιβλιοθηκών (ΔΙ.ΝΑΥ.ΒΙ.)
Maritime Heritage (MH)	Ναυτική/Ναυτιλιακή Κληρονομιά (NNK)
material objects	υλικά αντικείμενα
metal industry	βιομηχανία μετάλλου
NAFKLIROS (=Bosun)	ΝΑΥΚΛΗΡΟΣ
narrower term	ειδικότερος όρος
orphans	ορφανοί όροι
related term	σχετικός όρος
science	επιστήμη
top term	όρος κορυφής

Eleni Georgaki

PhD candidate

Department of Shipping, Trade and Transport

University of the Aegean

E-mail: egeorgaki@aegean.gr

Sofia Stamou

Assistant Professor (tenured)

Department of Archives

Library science and Museology, Ionian University

E-mail: stamou@ionio.gr