

Einsatz standardisierter Thesauri für Begriffe mit Bezug zum Meer und den Küsten

Christian Seip, Tillmann Lübker, Rainer Lehfeldt, Peter Korduan und Ralf Bill

Zusammenfassung

Bei der Integration vorhandener technischer Lösungen in die MDI-DE geht es nicht nur um die Einbeziehung der Metadaten, sondern auch darum diese zu verschlagworten und die Suche nach ihnen zu verbessern. Dazu eignen sich Schlagwortlisten, also kontrollierte Vokabulare und Thesauri. Beispiele für marine Thesauri sind „Küste“ und „NOKIS“ der BAW und der meeresökologische Thesaurus des BfN. Diese lagen zunächst jedoch in Tabellenstrukturen (Excel) vor und konnten ihre semantische Kraft dadurch nicht voll ausspielen. Übergibt man solche Wortlisten SKOS Management Werkzeugen wie iQvoc, lassen sie sich in vollem Umfang verwerten. Doch müssen die Wortlisten dafür zunächst in das SKOS Format überführt werden. Dieser Beitrag zeigt, wie die Überführung mit freien und Open Source Java Bibliotheken und Frameworks möglich ist, und zeigt abschließend darauf aufbauende Nutzungen der marinen Thesauri auf.

Schlagwörter

Thesaurus, Wortliste, Schlagwort, Meer, Küste, SKOS, Metadaten, iQvoc

Summary

When integrating technical solutions into MDI-DE it is not just about the inclusion of metadata, but also about keyword indexing these and to improve the search for these. For this purpose, word lists, i.e. controlled vocabularies and thesauri, can be used. Examples of marine thesauri are “Küste” and “NOKIS” maintained at BAW and the marine ecological Thesaurus implemented at BfN. Initially, these were available in table structures (Excel) and thus their semantic power could not be fully exploited. Passing such word lists to SKOS management tools such as iQvoc they can be used to their full extent. But the word lists have to be converted into the SKOS format first of all. This paper outlines these conversion options using free and open source Java libraries and frameworks and shows usages that build upon the conversion of the marine thesauri.

Keywords

thesaurus, word list, keyword, sea, coast, SKOS, metadata, iQvoc

Inhalt

1	Einführung	68
2	Vorhandene marine Thesauri und Schlagwortlisten	69
2.1	Küste und NOKIS.....	69
2.2	Meeresökologischer Thesaurus des BfN.....	70
3	Überführung in einen standardisierten und webbasierten Thesaurus.....	71
3.1	Technische Grundlagen (SKOS und iQvoc).....	71
3.2	Konzeption der Implementierung	73
3.3	Anwendung JSKOSify	73
4	Nutzung von standardisierten Thesauri.....	75
4.1	Aktualisierung und Pflege des Wortgutes	75
4.2	Bereitstellung nach innen und außen.....	76
4.3	Erfassung von Metadaten.....	77
4.4	Suche nach Daten	78
4.5	Parallele Nutzung standardisierter Thesauri.....	78
5	Zusammenfassung und Ausblick	79
6	Schriftenverzeichnis.....	80

1 Einführung

Als „Thesaurus“ bezeichnet der internationale Standard für Thesauri (ISO 25964) einen kontrollierten und standardisierten Wortschatz („vocabulary“), in dem als diskret wahrgenommenes – Objekte, Dinge oder abstrakte Einheiten – („*concepts*“) durch Begriffe („*terms*“) repräsentiert und beschrieben werden. Dabei können die Beziehungen der *concepts* zueinander abgebildet (z. B. durch Ober- und Unterbegriffe) und durch Normbegriffe („*preferred terms*“), Synonyme und Quasisynonyme ergänzt werden. Für den Bereich der Meere und Küsten existieren im deutschsprachigen Raum verschiedene Wortschatzsammlungen unterschiedlicher Ausprägungen und mit verschiedenen thematischen Schwerpunkten. Hierzu gehören das Stichwortverzeichnis Küste des Kuratoriums für Forschung im Küsteningenieurwesen (KFKI) sowie die NOKIS-Wortliste (s. Kapitel 2.1) für den Bereich der deutschen Küstenzone und der meeresökologische Thesaurus des Bundesamts für Naturschutz (BfN) (s. Kapitel 2.2) als Mikrothesaurus für den Themenbereich Meeresnaturschutz. Zudem sind weitere Wortschatzsammlungen für die Arbeit im marinen Kontext relevant, die über den Bereich der Küste und des Meeres weit hinausgehen, wie der Umwelt-Thesaurus (UMTHES) des Umweltbundesamtes (<http://data.uba.de>) und der General Multilingual Environmental Thesaurus (GEMET) der Europäischen Umweltagentur (<http://www.eionet.europa.eu/gemet>).

In Geodateninfrastrukturen wie der Marinen Daten-Infrastruktur Deutschland (MDI-DE) fällt dem Einsatz von standardisierten Thesauri eine besondere Bedeutung zu. Da verteilt vorliegende Daten und Dienste über die Metadaten auffindbar gemacht werden (WOSNIOK et al. 2014), ist es wichtig, dass für die Verschlagwortung in den Metadaten einheitliche Schlagwörter verwendet werden. Die Blesssralle (hier als Normbegriff

angesehen) beispielsweise gibt es auch in der Schreibweise „Blässralle“ und sie ist darüber hinaus auch als „Blässhuhn“ bzw. „Blesshuhn“ bekannt. Setzt man einen Thesaurus auch bei der Suche ein, so können auch Daten gefunden werden, die mit synonymen Begriffen, der wissenschaftlichen Bezeichnung oder dem englischen Namen verschlagwortet sind und ansonsten nicht auffindbar wären. Mit Hilfe eines multilingualen Thesaurus würde im obigen Beispiel auch eine Suche nach „*Fulica atra*“ oder „Eurasian Coot“ erfolgreich verlaufen. Ein relational aufgebauter Thesaurus ist darüber hinaus hilfreich, um eine Suche nach Daten einzugrenzen (durch Auswahl eines Unterbegriffes) oder zu erweitern (durch Auswahl eines Oberbegriffes, z. B. Rallenvögel). Zudem kann eine Abbildung des Thesaurus in Form einer hierarchischen Baumstruktur sehr hilfreich sein, geeignete Schlüsselwörter überhaupt erst zu finden.

Im Rahmen der MDI-DE wurde der Einsatz von Thesauri mit Bezug zum Meer und der Küste daher als wichtige Grundlage angesehen. Hauptziel war es dabei, die vorhandenen Wortschätze und Thesauri zu integrieren und allen Beteiligten auf einfache Weise zugänglich zu machen. Hierzu wurde eine technische Lösung favorisiert, bei der ein Thesaurus in ein standardisiertes Modell überführt wird (s. Abschnitt 3 in diesem Beitrag). Webbasierte Werkzeuge ermöglichen es dann, Anfragen an den Thesaurus zu stellen und diesen zu visualisieren sowie zu pflegen. Die auf diese Weise standardisierten Thesauri können durch die Projektpartner genutzt und im Portal der MDI-DE zusammengeführt werden (s. Abschnitt 4 in diesem Beitrag).

2 Vorhandene marine Thesauri und Schlagwortlisten

Neben dem bereits eingangs erwähnten Thesaurus UMTHEs auf nationaler Ebene und dem in engen Zusammenhang mit INSPIRE stehenden Thesaurus GEMET auf europäischer Ebene ist das Global Change Master-Directory von der NASA (GCMD, <http://gcmd.gsfc.nasa.gov>) auf weltweiter Ebene angesiedelt. Alle drei enthalten auch marine Begriffe, sind aber nicht Küsten- und Meeres-spezifisch. Die nachfolgend vorgestellten drei Thesauri auf nationaler Ebene (Küste, NOKIS und der meeresökologische Thesaurus des BfN) hingegen sind im Bereich Meer und Küste angesiedelt.

Darüber hinaus gibt es auch Ansätze, Thesauri zu verorten. Insbesondere bei Ortsnamensthesauri macht dies Sinn, wie der Deutsche Küstengazetteer unter Beweis stellt (Kohlus et al. 2014).

2.1 Küste und NOKIS

Die Wortliste „Küste“ vom Kuratorium für Forschung im Küsteningenieurwesen (KFKI) listet in einer Spalte deutsche Begriffe und in einer weiteren Spalte die englische Übersetzung von insgesamt 167 Begriffen. Bei den deutschen Begriffen findet man teilweise auch Verknüpfungen wie z. B. „Küstenabbruch s. Abrasion“.

Im Jahr 1983 erschien eine annotierte Bibliographie aller Beiträge, die in der Zeitschrift „Die Westküste“ von 1938 bis 1943 und „Die Küste“ von 1952 bis einschließlich Heft 37, 1982, veröffentlicht wurden (Bertram-Lyko 1983). Sie umfasste für die damaligen 395 Artikel auch ein Schlagwortverzeichnis und ein geographisches Verzeichnis. Diese beiden Listen werden mit jeder Ausgabe der Küste und des Newsletters KFKI aktuell in Deutsch und Englisch fortgeschrieben. Die geographischen Namen gehen in

den Küstengazetteer (KOHLUS et al. 2014) ein, und die Schlagwörter werden im Thesaurus der MDI-DE verwendet.

Ein weiteres Beispiel für eine Wortliste mit Verknüpfungen ist die Wortliste der Metadatenbank Nord-Ostsee-Küsten-Informationssystem (NOKIS) mit fast 3000 Zeilen. Neben Verknüpfungen enthält diese umfangreiche Wortliste teilweise Begriffserläuterungen in Deutsch und/oder Englisch sowie Quellenangaben.

Im NOKIS Projekt (LEHFELDT et al. 2014) entstand diese Wortliste mit wasserbaulichen Begriffen aus der DIN 4049 Hydrologie: 1-Grundbegriffe, 2-Gewässerbeschaffenheit, 3-quantitative Hydrologie zusammen mit Fachbegriffen aus den Dokumentationen der KFKI Küstenforschungsprojekte. Diese Liste wird kontinuierlich aktualisiert. Dabei wird versucht, sämtliche Begriffe bibliographisch bezüglich ihrer Semantik nachzuweisen und geeignete Übersetzungen mit Begriffsdefinitionen wie z. B. vom US Army Corps of Engineers (<http://chl.erdc.usace.army.mil/glossary>) anzugeben. Eine wesentliche Quelle für diese Arbeiten ist die mehrsprachige Übersetzung von Fachbegriffen der UNESCO (1998) sowie der GEMET und das Global Change Masterdirectory (OLSEN et al. 2013).

2.2 Meeresökologischer Thesaurus des BfN

Der meeresökologische Thesaurus des BfN ist ein Mikrothesaurus für den Themenbereich Meeresnaturschutz. Der Thesaurus wurde 2007 erstellt und umfasst ca. 2700 Begriffe und mehr als 3000 Synonyme. Neben alternativen Schreibweisen und synonymen Bezeichnungen enthält der Wortschatz, vor allem für die marinen Arten, die wissenschaftliche Bezeichnung sowie eine englische Übersetzung, die auch Synonyme mit einschließt.

Der meeresökologische Thesaurus ist hierarchisch gegliedert und umfasst 15 Einträge auf der obersten Ebene (vgl. Abb. 5 in Abschnitt 4). Einen Hauptschwerpunkt mit etwa einem Drittel aller Einträge bilden dabei die Bereiche „Marine Arten“ und „Biologie und Ökologie“. Neben weiteren meeresökologischen Kernthemen wie marine Biotoptypen oder Meereszonen werden durch den Thesaurus ebenfalls Themenkomplexe abgedeckt, die für die Arbeit von Behörden wichtig sind, wie Organisationen und Institutionen, Rechtsnormen und Begriffe zur Datenerhebung und Erfassungsmethoden. Die hierarchische Gliederung von Unterbegriffen erfolgt bis in die siebte Ebene wie am Beispiel des Schweinswales deutlich wird (Marine Arten -> Tierarten -> Säugetiere -> Wale -> Zahnwale -> Schweinswale -> Schweinswal). Der meeresökologische Thesaurus kann daher als strukturierte Wortsammlung des Fachgebietes angesehen werden.

Zunächst lag der Wortschatz in Form von strukturierten und die Hierarchien abbildenden Excel-Tabellen vor. Da diese nur schwer handhabbar sind, wurden die Informationen in eine Oracle-Datenbank überführt und in das Landschafts- und Naturschutzinformationssystem (LANIS) des BfN integriert. Dort wurde der Thesaurus für die Suche nach und die Verschlagwortung von Objekten verwendet. Für die Einbindung wurde jedoch kein standardisiertes Modell verwendet und ein Redaktionssystem war nicht vorhanden.

In enger Abstimmung mit den anderen Partnern der MDI-DE hat das BfN beschlossen, die im meeresökologischen Thesaurus enthaltenen Informationen zukünftig über eine offene Schnittstelle bereitzustellen. Hierzu wurde im Rahmen der MDI-DE ein Tool entwickelt, mit Hilfe dessen die in Excel vorliegenden Daten in das W3C-Standardformat Simple Knowledge Organization System (SKOS, <http://www.w3.org/2004/02/skos/>)

überführt werden konnten (s. Kapitel 3.2 und 3.3). Über eine Instanz der Software iQvoc können die Partner der MDI-DE sowie die Allgemeinheit den meeresökologischen Thesaurus nun einfach in standardkonformer Form abfragen. Zudem erleichtert dieser Weg die Aktualisierung und Pflege des Wortgutes. Im Rahmen der Weiterentwicklung der hausinternen Geodateninfrastruktur wird die iQvoc-Instanz weiterhin in die neuen Komponenten für Metadaten und für die Darstellung via WebGIS integriert. Diese Integration erleichtert daher auch die Erfassung von Metadaten und die Suche nach Daten.

3 Überführung in einen standardisierten und webbasierten Thesaurus

Möchte man die bestehenden Wortlisten z. B. für die Verschlagwortung oder Suche nutzen, müssen sie zunächst erst in ein spezielles Format überführt werden. Denn auch wenn ein Mensch vieles aus den Listen herauslesen kann, ist einem Computersystem dies nicht ohne weiteres möglich.

3.1 Technische Grundlagen (SKOS und iQvoc)

Das Simple Knowledge Organisation System (SKOS) ist ein Modell, um die Struktur und den Inhalt von Begriffssammlungen (SKOS spricht hier von „*concept schemes*“) zu repräsentieren und ist seit 2009 ein W3C-Standard (World Wide Web Consortium, <http://www.w3.org/2004/02/skos/>). Begriffssammlungen können dabei Thesauri, Taxonomien, Terminologien, Glossare, Wortlisten und andere kontrollierte Vokabulare sein. SKOS basiert auf dem Resource Description Framework (RDF) (MILES et al. 2005).

Wortlisten im SKOS-Format können SKOS-Managementwerkzeugen wie TopBraid Enterprise Vocabulary Net oder iQvoc übergeben werden, um sie im Browser zu visualisieren (z. B. in einer hierarchischen Ansicht), und zu pflegen. Dies ermöglicht es Fachleuten auf einfache Weise ihr Expertenwissen in solch ein kontrolliertes Vokabular einfließen zu lassen. SKOS ermöglicht verschiedene Beziehungen zwischen *concepts* wie *skos:broadener* und *skos:narrower*. Hiermit lässt sich beispielsweise spezifizieren, dass „Rotalgen“ zu den „Algen“ gehören (*skos:broadener*) und dass zum „Strand“ das untergeordnete *concept* „Strandabschnitt“ gehört (*skos:narrower*). Daneben gibt es die SKOS-Eigenschaft *skos:related* für *concepts*, die in irgendeiner Verbindung zueinander stehen. Um Verbindungen zwischen *concepts* bestehender Thesauri abzubilden, gibt es die SKOS-Eigenschaften *skos:exactMatch*, *skos:closeMatch*, *skos:broadMatch*, *skos:narrowMatch* und *skos:relatedMatch*, die ähnlich wie die Verbindungstypen innerhalb eines kontrollierten Vokabulars funktionieren. Kommt das *concept* „Strand“ mit der gleichen Bedeutung beispielsweise in zwei Thesauri vor, würde ein *skos:exactMatch* Verwendung finden (MILES und BECHHOFFER 2009).

Da die Ergebnisse der MDI-DE über ein Webportal (www.mdi-de.org) der Öffentlichkeit zugänglich gemacht werden, sollte auch für die marinen Wortschatzsammlungen ein webbasiertes SKOS-Managementwerkzeug zum Einsatz kommen. Nach Abwägung der Alternativen fiel die Entscheidung für iQvoc (innoQ Deutschland GmbH) als SKOS-Managementsystem. iQvoc wurde nach dem Model-View-Controller (MVC) Prinzip konzipiert. MVC ist ein Muster zur Strukturierung, das Software in die drei Einheiten Datenmodell (engl. *model*), Präsentation (engl. *view*) und Programmsteuerung (engl. *controller*) gliedert und somit voneinander trennt. Zudem bietet iQvoc eine mehrsprachige

Benutzeroberfläche mit einer Baumansicht der Konzepte, Suche usw. sowie eine Administrationsoberfläche u.a. mit Editierungsfunktion und einem feingranularen Redaktionssystem (Sperrung, Versionierung, Status, Konsistenzprüfung) (BANDHOLTZ et al. 2010).

iQvoc baut - wie auch SKOS - auf dem Linked-Data Konzept auf, dessen vier Prinzipien wie folgt lauten (BERNERS-LEE 2009):

- Verwende URIs als Namen für Dinge.
- Verwende HTTP URIs, damit diese Namen nachgeschlagen werden können.
- Wenn jemand einen Namen nachschlägt, biete nützliche Information und verwende dabei die Standards (RDF, SPARQL).
- Biete von dort auch Links auf andere URIs, damit man mehr Dinge entdecken kann.

MareThes

Konzepte

Kollektionen

Konzepte

Hierarchisch

LHM

Auswertung und Analyse

Biologie und Ökologie

Datenbanken

Datenerhebung und Erfassungsmethoden

Gefährdungsfaktoren und Auswirkungen

Marine Arten

Pflanzenarten

Tierarten

Amphibien

Fische und Rundmäuler

Knorpelfische

Adlerrochen

Alligatorhai

Blauhai

Blonde

Chagrin-Rochen

Dornhai

Eishai

Fleckhai

Fleckrochen

Fuchshai

Fyllas-Rochen

MareThes

Konzepte

Kollektionen

Suche

Über

Adlerrochen

Konzept

Basisdaten

Bezeichner

Beziehungen

Anmerkungen

Bevorzugte Labels

Deutsch

Adlerrochen

Bevorzugte Labels

Englisch

Spotted Eagle Ray

Alternative Labels

Englisch

Alternative Labels

Deutsch

Eulenrochen, Meeradler, Myliobatis aquila (Linnaeus, 1758)

Abbildung 1: a) Hierarchien in iQvoc b) Konzept „Adlerrochen“ in iQvoc.

Die eingangs beschriebenen Eigenschaften von *concepts* in Wortschatzsammlungen werden natürlich auch von iQvoc abgebildet, wie Abb. 1 (b) - mit der Angabe von vielen Synonymen und Übersetzungen - zeigt. *Concepts* werden in der Datenbank von iQvoc in der Tabelle *concepts* gespeichert und die eigentlichen Begriffe in der Tabelle *labels*, die n:m-Beziehung zwischen Konzepten und Begriffen in der Beziehungstabelle *labelings*. Die Beziehungen zwischen den *concepts* untereinander werden in der Tabelle *concept_relations* hinterlegt. Damit lassen sich Hierarchien innerhalb eines Thesaurus aufbauen, wie Abb. 1 (a) verdeutlicht. SKOS-Beziehungstypen wie *broadier* oder *narrower* werden in der Tabelle *matches* gespeichert. Weitere Tabellen dienen der Zuordnung von Begriffen zu Sammlungen (*collections*), Klassifizierungen

(`classifications` und `classifiers`) und Beschreibungen (`notes` und `note_annotations`). Die Tabelle `users` dient der Verwaltung von Rechten und Rollen.

3.2 Konzeption der Implementierung

Da die marinen Wortsammlungen der Projektpartner noch nicht in standardisierten Formen vorlagen, mussten diese zunächst in das SKOS-Format überführt werden. Bei kleineren Wortlisten wie beispielsweise der KFKI-Wortliste „Küste“ (167 *concepts*) wäre eine manuelle Umwandlung denkbar, aber spätestens bei umfangreichen Wortlisten wie der NOKIS-Wortliste mit knapp 3000 *concepts* würde dies kaum noch Sinn machen. Daher wurde ein Werkzeug entwickelt, das eine durch den Nutzer gesteuerte Umwandlung vornimmt.

Der erste Schritt zur Umwandlung ist das Einlesen einer Wortliste im Excel-Format. Anschließend müssen den Spalten, also z. B. die Spalte mit der englischen Übersetzung eines *concepts*, die SKOS-Eigenschaften und die Sprache, in der die Daten der Spalte vorliegen, zugeordnet werden. Da SKOS auf dem Linked-Data Konzept aufbaut, das URIs verlangt, benötigt man zur Sicherstellung einer eindeutigen Adresse für jedes *concept* einer jeden Wortschatzsammlung beim Aufbau des Dokumentes eine Basis HTTP URI.

Sobald diese Informationen vorliegen, können im Vorspann des RDF- bzw. SKOS-Dokuments die Namespaces entsprechend gesetzt werden. Anschließend werden die Wortlisten zeilenweise eingelesen und damit das SKOS-Dokument befüllt. Danach können Beziehungen - zwischen Begriffen innerhalb einer Wortliste oder zwischen verschiedenen Ontologien –gesetzt werden. Der letzte Schritt ist das Abspeichern der RDF-Struktur im XML-, Turtle- oder N-Triples-Format, wobei der iQvoc-Import zwingend das N-Triples-Format verlangt.

3.3 Anwendung JSKOSify

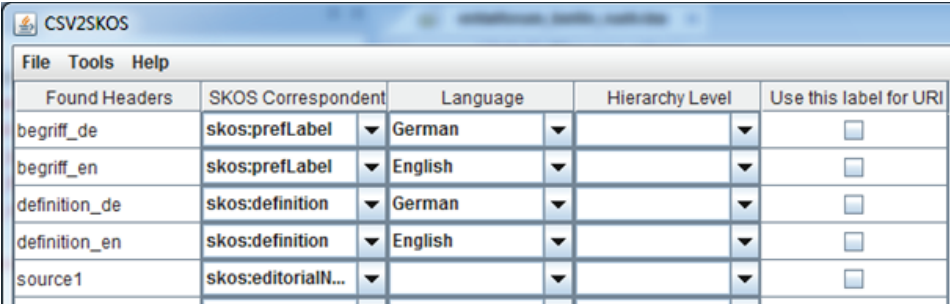
Aus der Konzeption ergaben sich folgende sechs Schritte, die für die Implementierung der Anwendung *JSKOSify* (<http://github.com/Sicky/JSKOSify>) umzusetzen sind:

1. Wortliste einlesen
2. Den Spalten die SKOS-Eigenschaften zuordnen
3. JDOM-Dokument aufsetzen
4. Dokument mit Begriffen/Konzepten füllen
5. Zusätzliche Dinge wie spezifischere Begriffe, Vorkommen eines Begriffes in anderen Ontologien usw. behandeln
6. Abspeichern der RDF-Struktur im XML-, Turtle- oder N-Triples-Format

Der erste Schritt – *das Einlesen der Wortliste* – wurde mit der freien Java Bibliothek CSV-Reader (<http://www.csvreader.com>) realisiert (Comma Separated Values [CSV] statt Excel-Format, da keinerlei Excel-spezifische Funktionalitäten in den Wortlisten genutzt werden). Einzige Voraussetzung für die zu verarbeitenden Wortlisten ist, dass ihre Spalten durch Spaltenüberschriften gekennzeichnet sind.

Wie Abb. 2 zeigt, erfolgt die *Zuordnung der SKOS-Eigenschaften zu den Spalten* (Schritt zwei) über eine Tabelle. In der ersten Spalte werden die gefundenen Spaltenüberschriften

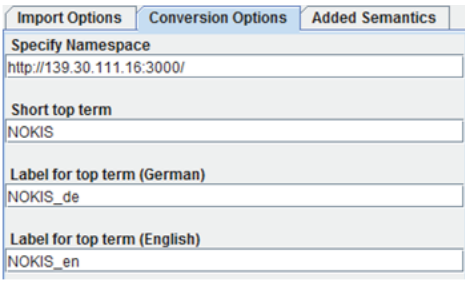
angezeigt und in den weiteren Spalten können dann u.a. SKOS-Eigenschaften (ob es sich beispielsweise um ein `skos:prefLabel`, eine `skos:definition` usw. handelt) und die Sprache den Spaltenüberschriften zugeordnet werden.



Found Headers	SKOS Correspondent	Language	Hierarchy Level	Use this label for URI
begriff_de	skos:prefLabel	German		☐
begriff_en	skos:prefLabel	English		☐
definition_de	skos:definition	German		☐
definition_en	skos:definition	English		☐
source1	skos:editorialN...			☐

Abbildung 2: Oberfläche zur Zuordnung der SKOS-Eigenschaften.

Im dritten Schritt wird das zu erzeugende *SKOS-Dokument konfiguriert*. Dazu werden ein Namespace und ein Wurzelement angegeben (siehe Abb. 3 (a)). Das Wurzelement vermeidet Uneindeutigkeiten beim mehrmaligen Auftauchen eines Begriffs (siehe 3.2) und ermöglicht es iQvoc darüber hinaus, alle Begriffe, die zu diesem Wurzelement (und damit zu dieser Wortliste) gehören, in der Baumansicht übersichtlich anzuzeigen (siehe Abb.3 (b)). Mit den Einstellungen, wie sie etwa in Abb. 3 (a) vorgenommen wurden, ist der Begriff „Strand“ unter der Adresse <http://139.30.111.16:3000/NOKIS#Strand> nachschlagbar.



Import Options **Conversion Options** Added Semantics

Specify Namespace

Short top term

Label for top term (German)

Label for top term (English)

Konzepte

 Hierarchisch

- ⊕ Kueste_de
 - ⊕ LHM
 - ⊖ NOKIS_de
 - Abbruchkante
 - Abdeckfundament

Abbildung 3: a) Oberfläche zur Spezifizierung der Basis URI usw. b) Wurzelemente in iQvoc.

Zur *Füllung des Dokuments mit Begriffen* (Schritt vier) wird die Wortliste zeilenweise durchlaufen und für jede Zeile ein neues *concept* angelegt. Danach werden verschiedene Kindelemente für das *concept* erzeugt, die dann die eigentlichen Informationen, wie Definitionen usw. enthalten. Zum Abschluss wird das neue *concept* dem JDOM-Dokument hinzugefügt.

Nach der Konvertierung der Wortliste in das SKOS-Format können *Beziehungen* (Schritt fünf) realisiert werden. Es können sowohl Beziehungen innerhalb der Wortliste als auch Beziehungen zu anderen Thesauri abgebildet werden. Eine Beziehung innerhalb einer Wortliste ist beispielsweise die Zuordnung des *concepts* „Strandabschnitt“ zum übergeordneten *concept* „Strand“ (*narrower*). Dazu können verschiedene Teile eines *concepts* untersucht werden und bei Übereinstimmung mit einem anderen *concept* die entsprechende SKOS-Relation wie `skos:broder`, `skos:related` usw. automatisch gesetzt werden. Im

zweiten Fall (Beziehungen zu anderen Thesauri) kann es von Interesse sein, wie ein *concept* in anderen Thesauri beschrieben wird. Dazu wird eine Anfrage an einen anderen Thesaurus über dessen API gesandt und das Ergebnis bzw. die URI des Ergebnisses mittels SKOS-Eigenschaften wie *skos:exactMatch*, *skos:closeMatch* usw. im *concept* gespeichert.

Im abschließenden Schritt sechs (*Speicherung des Dokuments*) wird das JDOM-Dokument gespeichert. Unter Nutzung des Sesame Frameworks von openRDF.org kann die Ausgabe in jedem beliebigen RDF-Format erfolgen (RÜH 2014).

4 Nutzung von standardisierten Thesauri

Anhand von vier Anwendungsfällen soll am Beispiel des meeresökologischen Thesaurus im Folgenden gezeigt werden, auf welcher unterschiedlichen Weise ein standardisierter mariner Thesaurus genutzt werden kann und welche Vorteile hierdurch entstehen. Die vorgestellten Anwendungsfälle sind auch auf die marinen Thesauri mit anderen Themenschwerpunkten übertragbar und zeigen auf, welche Rolle der Einsatz von standardisierten Thesauri generell in der MDI-DE spielt.

4.1 Aktualisierung und Pflege des Wortgutes

In der Software iQvoc ist bereits ein feingliedriges Redaktionssystem integriert, das die Benutzerrollen Editor, Publisher, Administrator und Gast vorsieht. Eine Veröffentlichung von Änderungen erfolgt stets erst nach der Freigabe, zudem werden Änderungen im System nachvollzogen, sodass eine Historisierung von Einträgen möglich ist. Eine eingebaute Konsistenzprüfung soll Flüchtigkeitsfehler vermeiden.

MareThes

DashboardKonzepteKollektionenSucheÜber

SchweinswalKonzept

Instanzkopie wurde erfolgreich angelegt und für andere Benutzer gesperrt.

Eigenschaften

Wiedervorlage

2013-04-18

Abgelaufen am

Labels

Mehrere Werte können kommagetrennt eingegeben werden.

Bevorzugte labels

de

Schweinswal

Bevorzugte labels

en

Harbour Porpoise

Alternative labels

en

Common Porpoise

Alternative labels

de

Phocoena phocoena (Linnaeus, 1758)

Abbildung 4: Web-basierte Oberfläche zur Aktualisierung und Pflege des Wortgutes.

Mit Hilfe dieses Redaktionssystems soll der meeresökologische Thesaurus zukünftig aktualisiert und gepflegt werden (s. Abb. 4). Ein Vorteil dieser Lösung ist, dass es die verteilte Bearbeitung (z. B. auch durch Auftragnehmer) eines zentralen Datenbestandes erlaubt. Ein solches Redaktionssystem, das die verteilte Bearbeitung eines zentralen Datenbestandes erlaubt, führt nicht nur zu einer hohen Aktualität, da Änderungen in Echtzeit aktiviert werden können, sondern vermeidet auch Inkonsistenzen, die durch eine dezentrale Pflege entstehen könnten. Jedoch sind Mechanismen der Qualitätssicherung hierbei unerlässlich. Im Zuge der Pflege und Aktualisierung des Thesaurus sollen auch Begriffe aus anderen Thesauri wie UMTHEs und GEMET über die Relationen „close match“ und „exact match“ verlinkt werden.

4.2 Bereitstellung nach innen und außen

Die Standardisierung des meeresökologischen Thesaurus ist hilfreich, um Technikexperten den Zugriff auf die darin enthaltenen Informationen zu ermöglichen. An die offene Schnittstelle der iQvoc-Instanz können verschiedene, klar definierte Abfragen gestellt werden. Die in Formaten wie HTML, RDF/XML oder Turtle zurückgelieferten Ergebnisse können dann von der anfragenden Stelle bzw. Anwendung ausgewertet, weiterverarbeitet und dargestellt werden.



HABITAT MARE

aktiv für die marine Lebensvielfalt

MEERESÖKOLOGISCHER THESAURUS DES BfN [Kurzanleitung \(PDF\)](#)

Der meeresökologische Thesaurus des BfN umfasst 2706 Schlagworte und 3068 Synonyme. Die fachspezifischen Schlagworte sind auf die Belange des Küsten- und Meeresnaturschutz abgestimmt und liegen in bis zu 8 hierarchisch strukturierten Ebenen vor. In dieser Webanwendung ist der Thesaurus durch eine hierarchische Baumstruktur abgebildet und kann per Volltextsuche durchsucht werden.

Gemarkte Begriffe (leeren): Seekabel

5. Merkleiste

Nach einem Begriff suchen

1. Suchmaske

☐ auch in Synonymen suchen?

☐ nur den exakten Begriff suchen?

☐ nur nach der Thesaurus-ID suchen?

Thesaurus als Baum durchsuchen

2. Baumansicht

- Marine Arten
- Marine Biotoptypen
- Meeres-, Öko- und Klimazonen, Biome und Faunenregionen
- Biologie und Ökologie
- Naturschutz, Planung und Genehmigung
- Rechtsnormen und allgemeine Rechtsbegriffe
- Organisationen, Institutionen, Programme und Stiftungen
- Nutzungen (Eingriffe)
- Gefährdungsfaktoren und Auswirkungen
- Datenerhebung und Erfassungsmethoden
- Auswertung und Analyse
- Qualitätssicherung
- Wissenschaftliche Disziplinen
- Datenbanken
- Quellentypen

Details zu einem Eintrag

3. Eintragsdetails

Begriff

Kormorane →

Wiss. Name

Phalacrocoracidae →

Synonym (en)

Cormorants →

Oberbegriff

Vogel →

Unterbegriffe

Kormoran →

Thesaurus-ID

1463

Es wurden 2 Einträge gefunden

4. Suchergebnis

- Kormorane
- Kormoran

[Login](#) | letzte Aktualisierung: Mai 2012 | [Kontakt / Fragen / Anregungen](#)

Abbildung 5: Prototypische Umsetzung einer Thesaurus-Webanwendung.

Neben dieser sehr technischen Möglichkeit kann der standardisierte Thesaurus über eine graphische Oberfläche der allgemeinen Öffentlichkeit im Internet und allen Behördenmitarbeitern im Intranet frei zugänglich gemacht werden. Bei dieser Bereitstellung für den Endanwender stehen der einfache Zugriff und eine intuitive Bedienbarkeit im Vordergrund. Eine entsprechende Webanwendung wurde bereits prototypisch umgesetzt (s. Abb. 5). In einer durch Autovervollständigung unterstützten Suchmaske kann der Nutzer nach einem Begriff suchen und dabei festlegen, ob auch Wortteile und Synonyme gefunden werden sollen. Zudem erleichtert eine graphische Baumstruktur den Zugang zum vorgehaltenen Wortschatz. In der Ergebnisansicht werden neben dem Begriff selber auch Synonyme, Übersetzungen sowie Unter- und Oberbegriffe angezeigt; zudem öffnet sich der Baum an der entsprechenden Stelle.

4.3 Erfassung von Metadaten

Auch bei der Erfassung von Metadaten wird der Thesaurus in der standardisierten Form eingesetzt. Wie bereits eingangs erläutert, werden in Geodateninfrastrukturen die Daten und Dienste über die Metadaten auffindbar gemacht (vgl. WOSNIOK et al. 2014), weshalb die Verwendung einheitlicher Schlagworte in den Metadaten eine wichtige Rolle einnimmt. Die Baumstruktur des Thesaurus sowie eine Funktion zur Autovervollständigung können dabei helfen, geeignete Schlagworte zu finden.

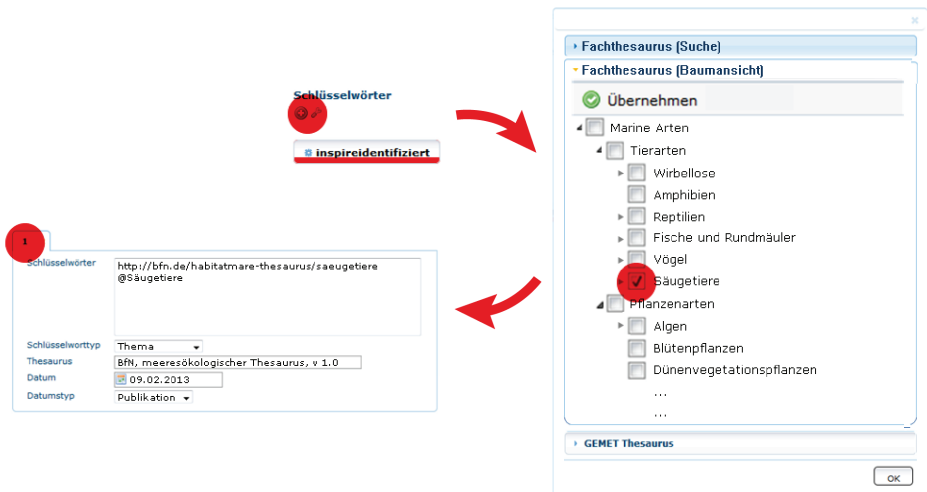


Abbildung 6: Integration standardisierter Thesauri in ein System zur Metadatenerfassung.

Der standardisiert vorliegende meeresökologische Thesaurus wurde in die Metadatenkomponente der internen Geodateninfrastruktur integriert. Hierzu wurde der im BfN verwendete Terra Catalog (<http://www.conterra.de/de/software/sdi>) um ein Dialogfenster erweitert (s. Abb. 6). Der Erfasser von Metadaten kann bei der Eingabe zwischen einem Suchfeld und einer Baumstruktur wählen. Außerdem können weitere, über eine standardisierte SKOS-Schnittstelle zugängliche, Thesauri wie UMTHEs und GEMET aufgerufen werden, ohne dabei das Dialogfenster zu verlassen. Nach Auswahl eines

Begriffes oder mehrerer Begriffe werden die dazugehörigen Informationen wie der Name des Thesaurus und die dem Begriff zugeordnete URI an die Metadatenkomponente übergeben.

4.4 Suche nach Daten

Nicht zuletzt unterstützt die standardisierte Form des Thesaurus beim Auffinden von Daten. In der Geodateninfrastruktur des BfN kann eine Suche nach Daten und Diensten an zwei verschiedenen Stellen erfolgen. Eine einfache Suche kann direkt aus der Darstellungskomponente (dem Intranet-WebGIS) heraus gestartet werden und zeigt die Treffer in der Metadatenkomponente an. Dort steht auch eine erweiterte Suche zur Verfügung mit Funktionen wie sie aus Metainformationssystemen bekannt sind. In der Ergebnisliste angezeigte Treffer können dann in die Darstellungskomponente übernommen werden. Beide Suchfunktionen nutzen die SKOS-Schnittstelle des meeresökologischen Thesaurus, sodass das einmal aufgebaute Vokabular auf vielfältige Weise verwendet werden kann. Dies vereinfacht u.a. auch das Auffinden von Daten, die mit Synonymen oder verwandten Begriffen verschlagwortet wurden. Analog zur Erfassung von Metadaten können auch bei der Suche weitere SKOS-basierte Thesauri angesprochen werden. Im Portal MDI-DE wurde derselbe Ansatz gewählt. Auch hier können SKOS-basierte Thesauri für die Suche in Metadaten verwendet werden.

4.5 Parallele Nutzung standardisierter Thesauri

Der gleichzeitige Einsatz verschiedener Thesauri, also z. B. den Thesaurus des KFKI für den Küstenbereich und den Mikrothesaurus des BfN für meeresökologische Themen, bietet verschiedene Vorteile. Jede Institution bzw. Behörde stellt das Spezialwissen der Experten auf diesem Gebiet bereit, womit eine hohe inhaltliche Qualität erreicht werden kann. Andere Institutionen, Behörden wie auch die Allgemeinheit können von diesem Spezialwissen profitieren und den systematisierten Wortschatz für ihre Zwecke verwenden, ohne dabei einen eigenen Wortschatz, z. B. für meeresökologische Themen, aufbauen und pflegen zu müssen. Die Standardisierung und die offene Bereitstellung über Web-Schnittstellen ermöglicht es dabei, verschiedene Thesauri miteinander zu kombinieren und gleichzeitig einzusetzen.

Bei der parallelen Nutzung verschiedener Thesauri kann es jedoch zu einem Problemfall kommen: Ein Begriff kann je nach Kontext unterschiedliche Bedeutungen haben, also ein anderes „Ding“ bezeichnen. So bezeichnet der Begriff Abundanz in der Ökologie die Anzahl an Individuen einer Art bezogen auf ihr Habitat (Dichte), in den Sprachwissenschaften hingegen die stilistische Überfüllung eines Ausdruckes (Redundanz, Pleonasmus). Folglich ist es richtig, bei der Verschlagwortung neben dem Begriff auch den Kontext mit aufzunehmen, um eine eindeutige Zuordnung zu gewährleisten. Dies kann z. B. in der folgenden Form geschehen:

- {<http://www.bfn.de/.../abundanz>}@{Abundanz}
- {<http://www.openthesaurus.de/synonyme/Abundanz>}@{Abundanz}

Diese eindeutige Form der Verschlagwortung führt jedoch dazu, dass bei einer kontextscharfen Suche auch nur diejenigen Metadaten gefunden werden, bei denen sowohl

Begriff als auch Kontext übereinstimmen; dies, auch wenn dasselbe „Ding“ gemeint ist. Folglich würde eine Suche nach dem Begriff „Säugetier“ beispielsweise im Kontext „BfN“ keine Treffer erzielen für Metadaten, die im Kontext „UMTHES“ erfasst wurden, denn:

- {<http://www.bfn.de/.../Säugetiere>}@{Säugetiere} ≠
- {http://data.uba.de/umt/_00021241}@{Säugetier} ≠
- {<http://www.eionet.europa.eu/gemet/concept/4982>}@{Säugetier}

Um diese ungewollten „Nicht-Treffer“ zu vermeiden, muss der bei der Suche in einem bestimmten Kontext ausgewählte Begriff in Begriff und Kontext zerlegt werden, so dass für die Suche der Kontext ignoriert werden kann. Hierbei muss die Applikationslogik entscheiden, wann eine Suche kontext-scharf durchgeführt werden soll und wann nicht.

In einem idealen System würden entweder dieselben „Dinge“ nicht doppelt in verschiedenen Thesauri vorgehalten werden oder aber es wären existierende Dubletten über match-Relationen miteinander verbunden. Da dies aber nur schwer realisierbar ist, ist der oben aufgezeigte Weg die praktikablere Lösung.

5 Zusammenfassung und Ausblick

Für den Einsatz von Thesauri mit Bezug zum Meer und der Küste in Geodateninfrastrukturen konnte im Rahmen des Projektes MDI-DE ein wichtiger Beitrag geleistet werden. Die Verwendung offener Schnittstellen wie RDF und XML sowie der Einsatz des Standardformates SKOS führen dazu, dass thematisch verschiedene Thesauri in eine einzelne Umgebung integriert und parallel genutzt werden können. Mit der hier vorgestellten Java-basierten Entwicklung *JSKOSify* können die Wortlisten, die zuvor in anderen Formaten vorlagen, komfortabel in das SKOS-Modell überführt werden.

Durch die Einbindung der standardisierten Wortschätze in die Software iQvoc können die Thesauri ohne weiteren Aufwand nach außen bzw. intern bereitgestellt und über ein Redaktionssystem gepflegt werden. Durch eine Integration in bestehende Systeme wie z.B. das Portal MDI-DE helfen die standardisierten Thesauri beim Auffinden von Daten und Diensten. Durch die Integration in interne Metainformationssysteme sind sie zusätzlich bei der Erfassung von Metadaten sehr nützlich.

Mit der Umsetzung des Stichwortverzeichnisses Küste des KFKI, der NOKIS-Wortliste sowie des meeresökologischen Thesaurus des BfN wurde ein guter Anfang gemacht für die Verwendung eines einheitlichen Vokabulars im Bereich der Küste und des Meeres. Weitere, bereits bestehende oder noch im Aufbau befindliche Wortlisten mit Bezug zum Meer und der Küste sollten entsprechend des vorgestellten Ansatzes folgen. Bei den Projektpartnern muss nun die technische Umsetzung und öffentliche Bereitstellung finalisiert werden. Zudem kommt es darauf an, die Thesauri fest in die Prozesse der täglichen Arbeit mit Geodaten zu integrieren. Dies betrifft neben der Suche nach Daten vor allem die Erfassung neuer und ggf. Aktualisierung bestehender Metadaten. Die hier vorgestellten Arbeiten sind zu verstehen als „best practice“-Beispiel für die Nutzung von standardisierten Thesauri in modernen Geodateninfrastrukturen.

6 Schriftenverzeichnis

- BANDHOLTZ, T.; SCHULTE-COERNE, T.; GLASER, R.; FOCK, J. und KELLER, T.: iQvoc – Open Source SKOS(XL) Maintenance and Publishing Tool. In: Proceedings of the Sixth Workshop on Scripting and Development for the Semantic Web, Crete, Greece, May 31, 2010.
- BERNERS-LEE, T.: Linked Data, 2009.
<http://www.w3.org/DesignIssues/LinkedData.html>
- BERTRAM-LYKO, A.: Bibliographie „Die Küste“. Die Küste, 39, 1983.
- ISO: NISO, 25964:2011/2013, Thesauri and interoperability with other vocabularies.
- KOHLUS, J.; SELLERHOFF, F.; VO, T.-T.-N.; LEHFELDT, R.; ROOSMANN, R. und ALCACER-LABRADOR, D.: Der Deutsche Küstengazetteer, ein service-basiertes Instrument zur Referenz und Kommunikation von Ortsbezeichnungen. Die Küste, 82, 2014.
- LEHFELDT, R.; REIMERS, H.-C. und KOHLUS, J.: NOKIS - Nord- und Ostseeküsten Informationssystem. Die Küste, 82, 2014.
- MILES, A.; MATTHEWS, B.; WILSON, M. and BRICKLEY, D.: SKOS core: Simple knowledge organisation for the Web. In: Proceedings of the 2005 International Conference on Dublin Core and metadata applications: Vocabularies in practice. Dublin Core Metadata Initiative, 2005.
- MILES, A. and BECHHOFFER, S.: SKOS Simple Knowledge Organization System Reference W3C, 2009.
<http://www.w3.org/TR/skos-reference>
- OLSEN, L.M.; MAJOR, G.; SHEIN, K.; SCIALDONE, J.; RITZ, S.; STEVENS, T.; MORAHAN, M.; ALEMAN, A.; VOGEL, R.; LEICESTER, S.; WEIR, H.; MEAUX, M.; GREBAS, S.; SOLOMON, C.; HOLLAND, M.; NORTHCUTT, T.; RESTREPO, R.A. and BILODEAU, R.: NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 8.0.0.0.0, 2013.
- RÜH, C.: Marine spatial data infrastructures – Approaches on evaluation, design and implementation. DGK Reihe C Heft 732. Verlag der Bayerischen Akademie der Wissenschaften, München 2014.
<http://www.dgk.badw.de/fileadmin/docs/c-732.pdf>
- UNESCO: International Glossary of Hydrology, Second Edition 1992, Version GB-DE, Deutsches Nationalkomitee für das Internationale Programm der UNESCO und das Operationelle Hydrologische Programm der WMO. IHP/OHP-Berichte, 12, 1998.
- WOSNIOK, C.; RÄDER, M.; KORDUAN, P. und LEHFELDT, R.: Metadaten in der MDI-DE. Die Küste, 82, 2014.