

Hochschule Hannover

Fakultät III - Medien, Information und Design

Abteilung - Information und Kommunikation

Vergleich von webbasierten Programmierschnittstellen zum Zugriff auf kontrollierte Vokabulare

Bachelorarbeit

im Studiengang Informationsmanagement

vorgelegt von

Moritz Horn

Erstgutachter: Prof. Dr. Christian Wartena

Zweitgutachter: Dr. Jakob Voß

Hannover, den 12.08.2016

Abstract

Diese Arbeit beschäftigt sich mit der Analyse von Webschnittstellen für kontrollierte Vokabulare von Informationseinrichtungen. Hierfür werden zunächst einige relevante Arten solcher Vokabulare vorgestellt. Außerdem wird ein Überblick über die in diesem Zusammenhang verwendeten Technologien und Standards für den maschinellen Datenaustausch über das World Wide Web gegeben. Darauf aufbauend werden schließlich die Schnittstellen von vier in der Bibliothekswelt viel genutzten Vokabularen miteinander verglichen. Zu diesem Zweck wird jedes Begriffssystem kurz vorgestellt und die Funktionalitäten der entsprechenden API untersucht. Danach werden die Untersuchungsobjekte nach ausgewählten Kriterien einander gegenübergestellt. Im Abschluss werden die Ergebnisse des Vergleichs präsentiert.

Inhaltsverzeichnis

Abkürzungsverzeichnis	I
Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
1 Einleitung	1
2 Kontrollierte Vokabulare	1
2.1 Klassifikationen	2
2.2 Thesauri	3
2.3 Normdateien	3
3 Webschnittstellen und -formate	4
3.1 RDF und Ontologien	5
3.1.1 RDFS und OWL	5
3.1.2 SKOS	7
3.1.3 Weitere etablierte Ontologien	7
3.2 RDF-Serialisierungen	8
3.2.1 RDF-XML	8
3.2.2 N-Triples	9
3.2.3 Turtle	9
3.2.4 JSON-LD	10
4 Untersuchte Schnittstellen	11
4.1 Regensburger Verbundklassifikation	12
4.2 Standardthesaurus Wirtschaft	18
4.3 Library of Congress Subject Headings	23
4.4 Gemeinsame Normdatei	25

5 Vergleich der Schnittstellen	27
5.1 Abfrageparameter und Dokumentation	28
5.2 Nutzung von Rückgabeformaten	30
5.3 Verwendete Ontologien	33
5.4 Interne Relationen	35
5.5 Verweise auf externe Ressourcen	36
6 Fazit	37
Eidesstattliche Erklärung	43

Abkürzungsverzeichnis

API	Application Programming Interface
DDC	Dewey Decimal Classification
DNB	Deutsche Nationalbibliothek
FoaF	Friend of a Friend
GND	Gemeinsame Normdatei
hbz	Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
IRI	Internationalized Resource Identifier
JS	JavaScript
JSON	JavaScript Object Notation
KOS	Knowledge Organisation System
LCCN	Library of Congress Control Number
LCSH	Library of Congress Subject Headings
LD	Linked Data
LoC	Library of Congress
MADS	Metadata Authority Description Schema
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
REST	Representational State Transfer
RVK	Regensburger Verbundklassifikation

SKOS	Simple Knowledge Organisation System
SPARQL	SPARQL Protocol and RDF Query Language
STW	Standardthesaurus Wirtschaft
Turtle	Terse RDF Triple Language
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
XML	Extensible Markup Language
ZBW	Zentralbibliothek für Wirtschaftswissenschaften

Abbildungsverzeichnis

4.1	Anzeige eines Konzeptes im RVK	14
4.2	Anzeige der Unterklassen eines Konzeptes im RVK	15
4.3	Anzeige der einem Konzept übergeordneten Klassen im RVK	16
4.4	Ergebnis der Begriffssuche im RVK	17
4.5	Rückgabe der „suggest“-Funktion des STW	19
4.6	Rückgabe der „suggest2“-Funktion des STW	19
4.7	JSON-Darstellung von Synonymen im STW	20
4.8	JSON-Serialisierung eines SKOS-Concepts im STW	21
4.9	Rückgabe der „mapping“-Funktion des STW	22
4.10	Turtle-Repräsentation eines STW-Deskriptors	23
4.11	SKOS-Repräsentation eines LCSH-Konzepts	25
4.12	GND-Eintrag in seiner JSON-Serialisierung in lobid	27
5.1	JSON-Serialisierung eines Konzepts der RVK-API	31
5.2	JSON-Serialisierung eines Mappings im STW	31
5.3	Verwendung der zbwext-Ontologie zum Verweis auf externen Gebrauch	34

Tabellenverzeichnis

4.1	Typen der Vokabulare und Betreiber der Schnittstellen	12
5.1	Dokumentation der APIs	28
5.2	Verfügbare Rückgabeformate der Schnittstellen	30
5.3	Verwendete RDF-Ontologien der Schnittstellen	33

1 Einleitung

Die Inhaltserschließung ist ein wichtiges Instrument für Informationseinrichtungen im Rahmen der Organisation und Präsentation von Medien und Informationen. Systeme wie Klassifikationen, Thesauri und Normdateien unterstützen in diesem Rahmen die Erstellung von konsistenten und qualitativ hochwertigen Erschließungsergebnissen (vgl. [KSS⁺14](#), S. 180f). Da die Erarbeitung solcher Systeme abhängig von der Komplexität des beschriebenen Themenbereichs sehr aufwändig sein kann, werden in der Praxis von vielen Einrichtungen oft etablierte Standards verwendet. Ein wichtiges Mittel für die Förderung des Austauschs solcher Systeme besteht in der Verwendung einer webbasierten Programmierschnittstelle, auf englisch Application Programming Interface (API). Für solche Schnittstellen wurden mittlerweile viele Standards entwickelt, die eine weite Verbreitung in allen Bereichen des Webs gefunden haben. Die Untersuchung dieses maschinellen Austauschs stellt den Rahmen dieser Arbeit dar. Sie steht außerdem im weiteren Zusammenhang mit Bestrebungen, Werkzeuge für die Verbesserung der Interoperabilität in diesem Gebiet zu entwickeln. Im ersten Abschnitt der Arbeit wird eine Übersicht über das Angebot von verschiedenen Typen kontrollierter Vokabulare, deren Struktur und Besonderheiten gegeben. Anschließend wird auf webbasierte APIs und die in diesem Zusammenhang für den Austausch relevanten Rückgabeformate und Metadatenmodelle eingegangen. Es folgt eine Vorstellung der untersuchten Vokabulare und der Funktionalitäten der Schnittstellen. Schließlich folgt der Vergleich der betrachteten Systeme und eine Zusammenfassung der Ergebnisse.

2 Kontrollierte Vokabulare

Kontrollierte Vokabulare definieren eine Sammlung an Begriffen, die Mehrdeutigkeiten in Form von Homonymen und Synonymen ausschließen und so von Methoden der Inhaltserschließung wie der freien Schlagwortvergabe abgegrenzt werden (vgl. [Spra](#)). Die breite Definition von kontrollierten Vokabularen wird durch verschiedene Unterkategorien weiter spezifiziert, um Daten strukturell an-

zuordnen und einen systematischen Zugriff auf spezielle Inhalte zu ermöglichen. Die entstehenden Begriffssysteme bieten sich auf Grund ihrer Eigenschaften unter anderem für die Verwendung durch Bibliotheken und anderen Informations-einrichtungen zur inhaltlichen Erschließung und damit einer Verbesserung der Suchbarkeit ihrer Bestände an. Im Englischen wird in diesem Zusammenhang von einem Knowledge Organisation System (KOS) gesprochen. Ein weiterer Schritt zum Ausbau solcher Systeme ist das Verknüpfen von einzelnen Begriffen mit deren Entsprechungen in anderen Vokabularen, auch Mapping genannt. Das Ergebnis des Mappings von mehreren (oder sogar allen) Begriffen zwischen zwei KOS wird als Konkordanz bezeichnet. Im Folgenden werden drei Typen von kontrollierten Vokabularen vorgestellt.

2.1 Klassifikationen

Eine Klassifikation bezeichnet ein kontrolliertes Vokabular, das durch den Prozess der Klassierung eine hierarchische Struktur zur Beschreibung eines Sachgebiets erzeugt. Die so entstandenen Klassen zeichnen sich in der Regel durch die Identifikation per Notationen aus, die als einzigartige Abfolgen von Buchstaben und/oder Zahlen nach unterschiedlichen Methoden gebildet werden. Zu unterscheiden sind präkombinierte Klassifikationen und Facettenklassifikationen. Während präkombinierte Klassen eine vollständige Aussage über die jeweilige Notation geben, repräsentiert eine Einfachklasse einer Facettenklassifikation lediglich ein bestimmtes Merkmal. Durch die Kombination mit anderen Einfachklassen kann in diesem Fall eine differenzierte Klasse erzeugt werden, die ebenfalls einen abgeschlossenen Bedeutungsrahmen widerspiegelt (vgl. [Ber05](#), Kap. 9).

Die Struktur einer Klassifikation wird durch verschiedene Relationen erzeugt, in denen Klassen miteinander stehen. Hierarchische Relationen dienen hierbei als Grundlage für die Über- und Unterordnung von Klassen zueinander. Klassifikationen sind Monohierarchien. Dies bedeutet, dass jede Klasse maximal eine übergeordnete Klasse haben kann, was die Bildung einer Baumstruktur bewirkt. Es werden zwei Methoden zur Hierarchiebildung unterschieden, die generische und die partitive. Die generische Relation verlangt, dass die untergeordnete Klasse neben den Merkmalen der Übergeordneten ein oder mehrere zusätzliche Eigenschaften besitzt. Dies findet beispielsweise in der Taxonomie von Lebewesen

statt. Im Gegensatz zu diesem Ansatz beschreibt die partitive Relation, dass eine Klasse einen Teil seiner übergeordneten Klasse darstellt. Große bibliothekarische Klassifikationen wie die Regensburger Verbundklassifikation (RVK) und Dewey Decimal Classification (DDC) verwenden das generische Relationsmodell.

2.2 Thesauri

Ähnlich der Klassifikation bildet ein Thesaurus eine Sammlung von Begriffen eines Themenbereichs und versieht diese untereinander mit Relationen. Wie die Klassifikation definiert der Thesaurus in diesem Rahmen zunächst ebenfalls hierarchische Relationen, um über- und untergeordnete Begriffe miteinander zu verbinden. Ergänzend dazu dient die Äquivalenzrelation der Darstellung von Begriffen, die miteinander bedeutungsgleich sind. Diese Begriffe werden in Äquivalenzklassen zusammengeführt und einer der Begriffe wird als vertretender Deskriptor bestimmt, während die restlichen als Nicht-Deskriptoren bezeichnet werden. So können etwa Synonyme verzeichnet und suchbar gemacht werden. Weiterhin definiert der Thesaurus Assoziationsrelationen, die zum Ausdruck bringen, dass zwei oder mehrere Begriffe sich teilweise überschneidende Bedeutungen haben (vgl. [Sprb](#)). Durch diese Prozesse entsteht so ein Gebilde mit bevorzugter Benennung und äquivalenten Begriffen, sowie den hierarchischen und assoziativen Relationen zu anderen Gebilden gleicher Art.

2.3 Normdateien

Normdateien sind Verzeichnisse, die im Informationswesen eine einheitliche Ansetzung und die eindeutige Identifizierung von Personen, Körperschaften und Begriffen fördern sollen. Die Dateien werden in der Regel durch übergeordnete staatliche Einrichtungen auf National- oder Sprachraumsebene gepflegt und bieten eine Möglichkeit für die Versorgung von Institutionen mit konsistenten und qualitativ hochwertigen Metadaten. In Deutschland wird die Gemeinsame Normdatei (GND) durch die Deutsche Nationalbibliothek (DNB) betrieben, wobei die Erschließung auch in Kooperation mit den Bibliotheksverbänden und einzelnen großen Bibliotheken betreut wird. Da der Rahmen von vielen Normdaten nicht nur auf Sachbegriffe beschränkt ist, benötigen sie eine größere Auswahl an Eigen-

schaften als Thesauri oder Klassifikationen. Beispielsweise werden mit Personen deutlich andere Informationen verknüpft als mit Körperschaften oder Orten. Der größte Anteil von Normdaten wird in der Regel durch solche Informationen zu Personen, Körperschaften und geographischen Orten ausgemacht.

3 Webschnittstellen und -formate

Der automatisierte Austausch von Informationen über das Web wird überwiegend durch HTTP-basierte APIs umgesetzt. Diese basieren auf einer an eine Datenbasis angeschlossenen Serverkomponente, die Anfragen eines bestimmten Formats entgegennimmt und eine Antwort basierend auf den gegebenen Parametern zurückgibt. Die für diese Arbeit relevante Art von Schnittstellen beruhen auf den Prinzipien des Representational State Transfer (REST), dem grundlegenden Verfahren für den Austausch von Daten über das World Wide Web. Das REST-Architekturmodell wurde 2000 durch Roy Fielding in seiner Dissertation spezifiziert und beschreibt Verhaltensregeln für die Server-Klienten-Kommunikation (vgl. [Fie00](#)). APIs, die auf REST basieren werden allgemein als „RESTful“ bezeichnet. Im Vergleich zum Datenaustausch über andere Netzwerkprotokolle wie zum Beispiel SOAP ¹ sind die Voraussetzungen für eine „RESTful API“ recht überschaubar. Wichtige Bedingungen im Bezug auf Webschnittstellen in diesem Kontext sind besonders die Verwendung von Uniform Resource Identifier (URI)s für die eindeutige Adressierbarkeit jeder verfügbaren Ressource, die Einheitlichkeit der Methoden (in der Regel die HTTP-Methoden GET, POST, PUT usw.) und die Entkopplung von Ressourcen und ihren Repräsentationen. Bei Letzteren sind neben der HTML-Ansicht etablierte Austauschformate wie Extensible Markup Language (XML) und JavaScript Object Notation (JSON) zu erwähnen.

Die Verwendung von Web-APIs spielt eine wichtige Rolle in der Realisierung des „Semantic Web“, einer Erweiterung des World Wide Web durch die Verknüpfung von Ressourcen. Dieses Ziel wird durch Empfehlungen des World Wide Web Consortium (W3C) für Standards bezüglich Datenformaten und Austauschprotokollen unterstützt. Eine der wichtigsten Empfehlungen in diesem Zusammenhang ist die des Resource Description Framework (RDF). Dieses Metadatenmodell wurde

¹Die neuste Version des Protokolls ist unter <https://www.w3.org/TR/soap12/> spezifiziert.

in direktem Hinblick auf die Bedürfnisse eines solchen semantischen Netzwerks kreiert und in Folge durch zahlreiche weitere Spezifikationen ergänzt. In diesem Abschnitt wird zunächst auf das Format an sich eingegangen, danach werden gebräuchliche Serialisierungen zur Representation dieser Daten erläutert.

3.1 RDF und Ontologien

Das RDF ist ein vom W3C herausgegebener Standard für die Beschreibung von Metadaten mit dem Zweck des Austauschs über das World Wide Web. Es ist ausgelegt auf den Datenaustausch zwischen Maschinen (also unter anderem APIs) mit einer Priorität auf die Erhaltung der Bedeutung der enthaltenen Daten ohne den Rückgriff auf externe Hilfsmittel. Das RDF-Modell basiert auf der Bildung von Aussagen durch die Verknüpfung von Ressourcen, die in einer vorgegebenen Anordnung stehen. Die drei Grundbestandteile eines solchen Statements werden als Subjekt, Objekt und Prädikat bezeichnet, ähnlich der Satzbildung im natürlichen Sprachgebrauch. Zusammen bilden die drei Satzteile ein „Triple“. Das Modell stellt zudem Anforderungen an die Form der einzelnen Bestandteile. Das Format sieht drei mögliche Arten von Ressourcen vor: URIs (bzw. deren internationalisierte Version IRI, „International Resource Identifier“), Literale (was einem Freitext entspricht) und „blank nodes“. Letztere sind dafür geeignet, ohne die Angabe einer URI auf die Existenz einer Ressource hinzuweisen. (vgl. zu diesem Absatz [W3C14](#)).

3.1.1 RDFS und OWL

Ein wichtiges Hilfsmittel für die Arbeit mit dem RDF-Modell stellt die Verwendung von Ontologien dar. Diese bestehen aus einer Menge an Aussagen über einen bestimmten Gegenstandsbereich mit Beziehungen zwischen deren Elementen ([HKPR12](#)). Sie werden in dem Datensatz durch eine URI referenziert, über dessen Adresse eine maschinenlesbare Spezifikation bezogen werden kann. Das Schema für die Definition von solchen so genannten *Namespaces* liefert zum einen das Resource Description Framework Schema (RDFS) und in Ergänzung dazu die Beschreibungssprache Web Ontology Language (OWL) ([BHBL09](#)).

Das RDFS ist eine W3C-Empfehlung, die ein Vokabular zur Beschreibung von Ontologien bereitstellt. Es besteht aus sechs Klassen und neun Eigenschaften, die wiederum ausschließlich auf das begrenzte RDF-Vokabular oder das RDFS selbst verweisen². Mit diesem Mittel werden zum Beispiel Klassen (`rdfs:Class`), Eigenschaften (`rdfs:Property`) und Verweise auf die URI der jeweiligen Ontologie (`rdfs:isDefinedBy`) bestimmt. Die Web Ontology Language erweitert dieses Modell um Expressionen für Einschränkungen und Beziehungen zwischen Klassen und Eigenschaften.

```
<owl:Class rdf:about="http://d-nb.info/standards/elementset/gnd#
  ↳ SubjectHeading">
  <rdfs:isDefinedBy rdf:resource="http://d-nb.info/standards/
    ↳ elementset/gnd#"/>
  <rdfs:label xml:lang="de">Schlagwort</rdfs:label>
  <rdfs:label xml:lang="en">Subject heading</rdfs:label>
  <rdfs:subClassOf rdf:resource="http://d-nb.info/standards/
    ↳ elementset/gnd#AuthorityResource"/>
  ...
  <owl:equivalentClass rdf:resource="http://www.w3.org/2004/02/skos/
    ↳ core#Concept"/>
</owl:Class>
```

In diesem Beispiel eines Ausschnitts der Spezifikation der GND Ontology³ sind einige dieser Funktionen dargestellt. Beschrieben wird die Klasse `gnd:SubjectHeading`, wie in dem OWL-Klassenparameter `rdf:about` bestimmt wird. Innerhalb dieser Klasse werden die Benennungen und die Hierarchie relation des Konzepts mit Hilfe von RDFS ausgezeichnet. Zusätzlich erfolgt ein Verweis durch `owl:equivalentClass` auf eine äquivalente Klasse in der SKOS-Ontologie.

²<https://www.w3.org/2000/01/rdf-schema>

³<http://d-nb.info/standards/elementset/gnd.rdf>

3.1.2 SKOS

Das für die Standardisierung im Web zuständige W3C hat seit August 2009 eine Empfehlung für die Kodierung von kontrollierten Vokabularen mit dem Namen Simple Knowledge Organisation System (SKOS). Diese formale Sprache hat das Ziel, Gemeinsamkeiten von verschiedenen Arten von kontrollierten Vokabularen zu identifizieren. So soll ein universelles Datenmodell für deren einfachen Austausch im Kontext des Semantic Web zu geschaffen werden (IS09). Auf Grund des Anspruchs der Interoperabilität ist SKOS von seinem Umfang her sehr begrenzt und verfügt nur über relativ generische Klassen. Die Ontologie verfügt über vier Klassen und 28 Eigenschaften⁴, mit denen ein beliebiges KOS in seinen Grundzügen abbildbar ist. Hierfür werden zunächst für das gesamte abzubildende Konstrukt die Klassen `skos:Collection`, `skos:ConceptScheme` und `skos:OrderedCollection` zur Verfügung gestellt. Das Hauptelement für die Beschreibung von einzelnen Entitäten kontrollierter Vokabulare wird durch das *Konzept* (`skos:Concept`) repräsentiert. Das Konzept findet etwa Anwendung als Klasse einer Klassifikation oder als Thesaurusterm. Neben diesen Klassen beinhaltet das SKOS Eigenschaften, die den jeweiligen Klassen zugeordnet werden können. Beispielsweise können einem Konzept ein bevorzugter Name (`skos:prefLabel`) zugewiesen werden, eine hierarchische Relation zu einem über- oder untergeordneten Konzept (`skos:broader` bzw. `skos:narrower`) oder ein Verweis auf sein `skos:ConceptScheme` durch die Eigenschaft `skos:inScheme`.

3.1.3 Weitere etablierte Ontologien

Neben dem für die Erschließung von kontrollierten Vokabularen sehr verbreiteten SKOS gibt es zahlreiche andere viel verwendete Ontologien. Diese variieren von sehr engen Anwendungsbereichen mit wenigen Klassen bzw. Eigenschaften (z.B. VANN) bis hin zu hochkomplexen Systemen mit hunderten von Elementen (z.B. die GND Ontology der DNB). Einen guten Überblick über die externe Nutzung von Vokabularen gibt das Dataset „Linked Open Vocabularies“⁵, in dem zur Zeit 561 Vokabulare mit ihren Verbindungen untereinander verzeichnet sind.

⁴<https://www.w3.org/2009/08/skos-reference/skos.html>

⁵<http://lov.okfn.org/dataset/lov>

3.2 RDF-Serialisierungen

Da das Resource Description Framework an sich keine Auszeichnungssprache repräsentiert, sondern lediglich ein Modell für die Beschreibung von Ressourcen, werden für den Austausch von Daten in RDF Syntaxen benötigt, die diesen Bedarf erfüllen. In Folge dessen wurden für die meisten Austauschformate Standards entwickelt, die eine Serialisierung des RDF-Modells ermöglichen.

3.2.1 RDF-XML

RDF-XML war zur Zeit der Entstehung des RDF-Standards die einzige Serialisierung zur Darstellung des Formats (W3C14). Die Elemente der bereits etablierten Auszeichnungssprache XML wurden hierbei verwendet, um die RDF-Graphen abzubilden. Für die Abbildung von Ontologien, einschließlich des RDF-Namespaces werden zunächst deren URIs mit dem Namespace-Prefix `xmlns` festgelegt, wobei für RDF und RDFS die Kürzel `rdf` und `rdfs` reserviert sind. Die Zuweisung der Namespaces für das RDF-Vokabular selbst sowie des SKOS-Vokabulars in RDF/XML sieht beispielsweise wie folgt aus:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  ↪ xmlns:skos="http://www.w3.org/2004/02/skos/core#"></rdf>
```

Diese Zuweisungen wird in der Regel für das gesamte Objekt vorgenommen, kann jedoch auch lokal für einen einzelnen Knoten deklariert werden. Die Graphenstruktur von RDF wird in XML/RDF durch eine Schachtelung von Elementen erreicht. Hierbei wechseln sich Elemente für Knoten und Prädikate ab. Ein einfaches Beispiel für die Darstellung einer Relation von zwei Konzepten im Standardthesaurus Wirtschaft (STW) lässt sich wie folgt darstellen:

```
<rdf:Description rdf:about="http://zbw.eu/stw/descriptor/18250-1">
  <skos:broader rdf:resource="http://zbw.eu/stw/descriptor
  ↪ /18353-5"/>
</rdf:Description>
```

In diesem Beispiel wird das Tag `rdf:Description` in Verbindung mit dem Attribut `rdf:about` verwendet, um das Subjekt des Statements (die URI eines Konzepts)

zu definieren. Das Tag des untergeordneten Elements entspricht der URI des Prädikats (in diesem Fall die Aussage, dass ein übergeordneter Term folgt), und der Wert des folgenden Attributs `rdf:resource` bildet den Knoten des Objekts. Aus diesem Beispiel lässt sich bereits erkennen, dass die XML-Syntax nicht gut für die Darstellung von RDF-Graphen geeignet ist. Aus diesem Grund haben sich andere auf RDF basierende Formate deutlich stärker durchgesetzt.

3.2.2 N-Triples

N-Triples ist eine Serialisierung für RDF, die ursprünglich lediglich als Testumgebung gedacht gewesen war (Bec14). Sie definiert Zeilenbegrenzte „Triple“, die jeweils ein RDF-Statement enthalten. Das Format verfügt über eine sehr begrenzte Anzahl an Syntaxregeln für die Darstellung der verschiedenen Ressourcentypen. Es wird festgelegt, dass Internationalized Resource Identifier (IRI)s in spitzen Klammern stehen und Literale mit Anführungszeichen versehen werden müssen. Für Letztere dient der Anhang eines „@“ als Einleitung für ein Sprachkürzel bzw. die Zeichenfolge „^^“ zum Anhang einer URI eines Datentyps. Die einzelnen Terme werden durch ein Leerzeichen getrennt und Statements werden mit „.“ abgeschlossen.

```
<http://id.loc.gov/authorities/subjects/sh85052220> <http://www.w3.  
  ↪ org/2004/02/skos/core#prefLabel> "Fuel cells"@en .
```

Dieses Beispiel zeigt die Zuweisung eines bevorzugten Terms (`skos:prefLabel`) in Form eines Literals mit einer Sprachbezeichnung zu der URI eines Normdatensatzes der Library of Congress Subject Headings (LCSH).

3.2.3 Turtle

Die Turtle-Serialisierung für das RDF bildet eine Obermenge des zuvor beschriebenen N-Triple-Formats. Zunächst ergänzt sie dieses Modell um eine Möglichkeit, ähnlich wie in RDF/XML Namespaces zu definieren. Ziel davon ist es, das Dokument übersichtlicher zu gestalten. Dieser Vorgang wird am Anfang des Dokuments durchgeführt und basiert auf der Verwendung zweier Direktiven. Mit dem Präfix `@base` wird die Standardumgebung des Dokuments festgelegt, innerhalb dessen

relative Pfade zu Ressourcen interpretiert werden. Die Direktive `@prefix` weist einem mit „:“ abgeschlossenen Kürzel die URI einer Ontologie zu, woraufhin das Kürzel als Vertreter der Ontologie zusammen mit deren Klassen und Eigenschaften verwendet werden kann:

```
@base <http://zbw.eu/stw/> .
```

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
```

```
<descriptor/18250-1>
```

```
  a skos:Concept;
```

```
  skos:prefLabel "Brennstoffzelle"@de, "Fuel cell"@en .
```

Das Beispiel zeigt einen Ausschnitt aus der Rückgabe eines Konzepts des STW. Im Kopf der Rückgabe wird der API-Endpunkt als Basis-URL definiert, sowie die Verwendung des Kürzels `skos:` für die Verwendung von SKOS-Elementen. Anschließend wird dem Konzept (repräsentiert durch eine gekürzte URI) die SKOS-Klasse `skos:Concept` zugewiesen, sowie zwei Labels in unterschiedlichen Sprachen. Dies entspricht einem RDF-Graphen mit zwei Statements.

3.2.4 JSON-LD

Eins der meistgenutzten Austauschformate im Web ist die JavaScript Object Notation. Sie beruht auf dem Objektschema von JavaScript, einem Standard für die Entwicklung von dynamischen Websites. Der Vorteil von JSON gegenüber anderen Formaten wie XML ist deshalb, dass JSON-Objekte mit wenig Aufwand in JavaScript genutzt werden können. JavaScript stellt hierfür einen nativen Parser bereit. Dies erleichtert unter anderem auch die Nutzung von JSON-basierten APIs in anderen Webseiten, die mit JavaScript arbeiten. Für die Serialisierung von RDF-Graphen wurde JSON-LD (wobei LD für „Linked Data“ steht) vom W3C als Standard entwickelt, wobei der Fokus auf einen möglichst geringen Aufwand für JS-Entwickler lag (vgl. [SLK⁺14](#)). Um einen Graphen darzustellen, werden für JSON-LD zunächst eine Reihe von *Keywords* definiert, die jeweils durch die Voranstellung eines „@“ gekennzeichnet werden. Zwei grundlegende Keywords hierfür sind `@id` und `@context`. Der Schlüssel `@id` deklariert hierbei die Verwendung einer URI innerhalb des RDF-Graphen. Unter dem Key `@context` werden,

wie in Turtle und XML/RDF, Namespaces definiert. Dies geschieht über eine einfache Key-Value-Zuweisung mit der entsprechenden URI als Wert. Im folgenden Ausschnitt eines JSON-LD-Objekts werden einige SKOS-Terme einzeln definiert:

```
"@context": {
  "prefLabel": "http://www.w3.org/2004/02/skos/core#prefLabel",
  "exactMatch": "http://www.w3.org/2004/02/skos/core#exactMatch",
  "closeMatch": "http://www.w3.org/2004/02/skos/core#closeMatch",
  "relatedMatch": "http://www.w3.org/2004/02/skos/core#
    ↪ relatedMatch",
}
```

Alternativ kann mit der für die Serialisierung von RDF üblichen Notation der Trennung von Namespaces und deren Eigenschaften durch einen Doppelpunkt gearbeitet werden:

```
{
  "@context": {
    "skos": "http://www.w3.org/2004/02/skos/core"
  },
  "skos:notation": "TK2931"
}
```

4 Untersuchte Schnittstellen

Im nächsten Abschnitt sollen eine Reihe von APIs zum Zugriff auf die Daten von etablierten kontrollierten Vokabularen miteinander verglichen werden. Solche Schnittstellen bieten die Möglichkeit, den Austausch solcher Informationen zu automatisieren, um sie zum Beispiel dynamisch auf Internetseiten anzeigen zu lassen. Die Auswahl dieser Systeme wurde im Hinblick darauf getroffen, ob sie eine überregionale Bedeutung haben und für sie eine öffentliche API zur Verfügung steht. In Tabelle 4.1 sind die ausgewählten Untersuchungsobjekte mit ihrem Typ und dem Betreiber der Schnittstelle dargestellt. Zu Beginn muss erwähnt werden, dass sowohl viele weniger verbreitete als auch eine Reihe sehr etablierter KOS (etwa die Dewey Decimal Classification) keine öffentlichen Schnittstellen zu ihren Metadaten unterhalten. Aus diesem Grund kann diese

Arbeit keine Vollständigkeit im Bereich der möglichen Umsetzungen von APIs im Bereich des Informationswesens gewähren.

Vokabular	Typ	API-Betreiber
Regensburger Verbundklassifikation	Klassifikation	Uni Regensburg
Library of Congress Subject Headings	Thesaurus	Library of Congress
Standardthesaurus Wirtschaft	Thesaurus	ZBW
Gemeinsame Normdatei	Normdatei	hbz

Tabelle 4.1: Typen der Vokabulare und Betreiber der Schnittstellen

In diesem Kapitel werden die untersuchten Vokabulare kurz vorgestellt und die Eigenschaften ihrer Schnittstellen genauer beschrieben. Zu diesem Zweck wurden sowohl die Onlinedokumentationen der jeweiligen APIs als auch die Rückgaben von Testabfragen ausgewertet. Als Rückgabeformat der in diesem Zusammenhang präsentierten Ausschnitte von zurückgelieferten Datensätzen wird JSON verwendet. Um die Vergleichbarkeit zwischen den einzelnen Schnittstellen zu ermöglichen wurde für alle Anfragen – außer anderweitig erwähnt – der (passendste) Datensatz des Begriffs „Brennstoffzelle“ bzw. des englischen Äquivalents „fuel cell“ verwendet.

4.1 Regensburger Verbundklassifikation

Die RVK wurde Anfang der 1980er Jahre von der Universität Regensburg erstmals schriftlich niedergelegt und wird laut den Betreibern derzeit von über 140 Bibliotheken für die Inhaltserschließung verwendet (vgl. Uni). Sie ist eine Universalklassifikation ohne Beschränkung auf einen thematischen oder geographischen Bereich, und besitzt 33 Oberklassen mit etwa 850000 Unterklassen. Diese Oberklassen werden hierbei durch eine Notation bestehend aus Großbuchstaben repräsentiert, auf der zweiten Hierarchieebene wird ein weiterer Buchstabe angehängt. Die Notationen aller weiteren Ebenen werden durch die Ergänzung von drei bis fünf Ziffern gebildet, die jedoch keine Rückschlüsse auf die jeweilige Hierarchieebene zulassen. Gebräuchlich sind auch durch Bindestriche verbundene Bereiche zwischen zwei Einzelnotationen, die auf einer Ebene eine eigene Notation bilden. Auf den darunter verzeichneten Ebenen sind dann Notationen innerhalb dieses Wertebereichs verzeichnet.

Beispiel: Antike Bibliotheken AN 59100 H694 (vgl. Wik15)

- A – Allgemeines
 - AN – Buch- und Bibliothekswesen, Informationswissenschaft
 - AN 50000 – AN 89900 Bibliothekswesen
 - AN 58000 – AN 64950 Biographie, Geschichte
 - AN 59000 – AN 59600 Bibliotheksgeschichte
 - AN 59100 Altertum
 - H694 – Verschlüsselung des Familiennamens – Hoepfner, Wolfram [Hrsg.]

Zusätzlich zu den Hierarchierelationen können RVK-Klassen ebenfalls über Registereinträge und Bemerkungen verfügen. Registereinträge sind hierbei Freitexteinträge, die den Kontext der jeweiligen Klasse erläutern. Die Bemerkungen enthalten Hinweise auf Klassen, die mit dem Konzept in Verbindung stehen.

Die 2010 eingeweihte RVK-API(HWN) stellt das KOS über eine offene Webschnittstelle zur Verfügung. Zur Zeit der Erstellung dieser Arbeit funktioniert die dokumentierte API der RVK nur eingeschränkt, es gibt jedoch eine zweite (undokumentierte) Version der Schnittstelle, in der weniger Fehler vorliegen. Der Aufbau einer Abfrage an die API sieht zunächst die Definition des Rückgabeformats vor. Anschließend wird der Typ der Abfrage bestimmt, worauf der gesuchte Begriff folgt. Eine vollständige Abfrage lässt sich wie folgt darstellen:

[http://rvk.uni-regensburg.de/api/\[Ausgabeformat\]/\[Abfragetyp\]/\[Suchbegriff\]](http://rvk.uni-regensburg.de/api/[Ausgabeformat]/[Abfragetyp]/[Suchbegriff])

Für die Verwendung des undokumentierten Endpunkts ist das „api“ durch „Portal_API“ zu ersetzen. Im Rahmen der Untersuchung wurde die undokumentierte Version verwendet. Für das Ausgabeformat stehen die Parameter XML und JSON zur Verfügung, im Folgenden werden die Rückgaben im JSON-Format für die Darstellung der Beispiele genutzt. Die Schnittstelle bietet fünf Abfragetypen an:

node

Der Abfrageparameter „node“ ermöglicht die Suche nach einem spezifischen Konzept per Notation. Dieser Endpunkt beschreibt dementsprechend eine grundlegende Funktion im maschinellen Retrieval. Trotz ihrer Bedeutung ist eine solche Abfrage jedoch nur in der undokumentierten API funktionsfähig. In Abbildung 4.1 ist das Resultat einer Beispielabfrage im JSON-Format dargestellt.

```
{
  "host": "rvk.uni-regensburg.de",
  "request": "/Portal_API/json/node/VN%206050%20-%20VN%206059",
  "node": {
    "notation": "VN 6050 – VN 6059",
    "benennung": "Brennstoffzellen; Brennstoffe Batterien;
      ↳ Angewandte Elektrochemie; Solarzellen; Wasserstoff als
      ↳ Energieträger",
    "bemerkung": "s.a. VE 6300 Elektrochemie; s.a. VE 6350
      ↳ Elektrolyte; s.a. VG 8000 Elektroanalyse; s.a. ZP 4120",
    "verweis": "",
    "has_children": "yes",
    "register": ["Batterie", "Brennstoff", "Brennstoffzelle", "
      ↳ Elektrochemie", "Energieträger", "Solarzelle", "Technik",
      ↳ "Wasserstoff"]
  },
  "root_name": "UB Regensburg"
}
```

Abbildung 4.1: Anzeige eines Konzeptes im RVK

Zu beobachten ist, dass die Anzeige der Klasse weder auf übergeordnete noch direkt auf untergeordnete Konzepte verweist, sondern lediglich durch einen Hinweis im Feld „has_children“ durch ein Literal „yes“ die Existenz von Unterklassen angezeigt. Um in diesem Fall die Verweise zu erhalten, müssen für das jeweilige Konzept weitere Abfragen mit den Typen „children“ und „ancestors“ durchgeführt werden.

children

Der zurückgelieferte Datensatz für den Abfrageparameter „children“ gibt eine Liste von direkt untergeordneten Klassen zurück (s. Abb 4.2), wobei die angezeigten Klassen in dieser Ansicht lediglich die Notationen und Terme, sowie das Feld `has_children` darstellen.

```
{
  "host": "rvk.uni-regensburg.de",
  "request": "/api/json/children/VN%206050%20-%20VN%206059",
  "node": {
    "children": {
      "node": [
        {
          "notation": "VN 6050",
          "benennung": "Allgemeines",
          "has_children": "no"
        },
        {
          "notation": "VN 6051",
          "benennung": "Lehrbücher",
          "has_children": "no"
        }
        ...
        {
          "notation": "VN 6059",
          "benennung": "Festschriften bei sachlichem Gesichtspunkt", "
            ↪ has_children": "no"
        }
      ]
    },
    "notation": "VN 6050 - VN 6059",
    "benennung": "Brennstoffzellen; Brennstoffe Batterien; Angewandte
      ↪ Elektrochemie; Solarzellen; Wasserstoff als Energieträger",
      ↪ "has_children": "yes"
  },
  "root_name": "UB Regensburg"
}
```

Abbildung 4.2: Anzeige der Unterklassen eines Konzeptes im RVK

ancestors

Der Parameter „ancestors“ gibt die gesamte dem angegebenen Knoten übergeordnete Hierarchie bis zu der allgemeinen Oberklasse zurück, der das Konzept untergeordnet ist. Diese Abfrage gibt ebenfalls die Klasse selbst wieder, jedoch sind hier nur die beiden Felder notation und benennung vorhanden. Die zurückgelieferte Struktur ist eine Verschachtelung mit der abgefragten Klasse als oberstes Element und der jeweils direkten Oberklassen als Unterobjekt (s. Abb. 4.3).

```
{
  "host": "rvk.uni-regensburg.de",
  "request": "/api/json/ancestors/VN%206050%20-%20VN%206059",
  "node": {
    "ancestor": {
      "node": {
        "ancestor": {
          "node": {
            "ancestor": {
              "node": {
                "notation": "V",
                "benennung": "Chemie und Pharmazie"
              }
            },
            "notation": "VN",
            "benennung": "Angewandte und technische Chemie"
          }
        },
        "notation": "VN 5000 – VN 6109",
        "benennung": "Angewandte und Technische Chemie allgemein"
      }
    },
    "notation": "VN 6050 – VN 6059",
    "benennung": "Brennstoffzellen; Brennstoffe Batterien; Angewandte
      ↳ Elektrochemie; Solarzellen; Wasserstoff als Energieträger
      ↳ "
  },
  "root_name": "UB Regensburg"
}
```

Abbildung 4.3: Anzeige der einem Konzept übergeordneten Klassen im RVK

nodes

Dieser Parameter führt eine Begriffssuche über die Benennungen und Register-
einträge aller Konzepte im RVK durch und gibt die gefundenen Klassen zurück (s.
Abb. 4.4).

```
{
  "host": "rvk.uni-regensburg.de",
  "request": "/Portal_API/json/nodes/Brennstoff",
  "node": [{
    "notation": "VN 6050 – VN 6059",
    "benennung": "Brennstoffzellen; Brennstoffe
      ↳ Batterien; Angewandte Elektrochemie;
      ↳ Solarzellen; Wasserstoff als Energieträger",
    "register": ["Batterie", "Brennstoff", "
      ↳ Brennstoffzelle", "Elektrochemie", "Energieträ
      ↳ ger", "Solarzelle", "Technik", "Wasserstoff"]
  }, {
    "notation": "ZK 6600",
    "benennung": "Brennstoffveredelung",
    "register": ["Brennstoff", "Veredelung"]
  }, {
    "notation": "ZN 8750",
    "benennung": "Brennstoffelemente"
  }, {
    "notation": "ZP 3270",
    "benennung": "Brennstoff- und Feuerungstechnik (
      ↳ Brenner; Feuerungen; Industrieöfen)",
    "register": ["Brennstoff", "Feuerung", "
      ↳ Industrieofen", "Technik"]
  }, {
    "notation": "ZP 4400",
    "benennung": "Energieeffizienz, Energieeinsparung,
      ↳ Energierückgewinnung",
    "register": ["Brennstoffeinsparung", "Elektrizität
      ↳ einsparung", "Energieeffizienz", "
      ↳ Energieeinsparung", "Energierückgewinnung", "
      ↳ Least Cost Planning"]
  }],
  "root_name": "UB Regensburg"
}
```

Abbildung 4.4: Ergebnis der Begriffssuche im RVK

register

Der Parameter führt laut der Dokumentation eine Begriffssuche über die Registerinträge aller Konzepte durch. In keiner der beiden API-Versionen liefert diese Funktion jedoch ein Ergebnis zurück.

4.2 Standardthesaurus Wirtschaft

Der STW der Zentralbibliothek für Wirtschaftswissenschaften (ZBW) hat einen Umfang von ca. 6000 Schlagwörtern und 20000 ergänzenden Synonymen (vgl. [Deu15](#)). Er vereinigt sieben Subthesauri verschiedener wirtschaftlicher und verwandter Themengebiete. Diese oberste Ebene der Subthesauri wird durch eine Notation bestehend aus einem einzelnen Großbuchstaben repräsentiert, alle weiteren Ebenen werden von der oberen getrennt durch einen Punkt durch zweistellige Zahlenkombinationen dargestellt. Die einzelnen Konzepte der STW-Systematik sind mit Deskriptoren verknüpft, die ebenfalls in einer hierarchischen Struktur untereinander stehen. Deskriptoren können wiederum mit mehreren Stellen in den Subthesauri verknüpft sein. Ebenfalls verfügen die Daten auch Referenzierungen zu äquivalenten Begriffen in anderen Systematiken, wie der GND und der DBPedia.

Die ZBW bietet zwei APIs für ihren Thesaurus an. Alle Daten des STW sind sowohl in Deutsch als auch auf Englisch verzeichnet. Zum einen gibt es die Möglichkeit, auf die einzelnen Konzepte des STW in mehreren Formaten zuzugreifen. Der Zugang hierzu ist über die Anzeige von Konzepten in der HTML-Anzeige ersichtlich. Die „Web Services for Economics“ sind im Rahmen der ZBW Labs zu erreichen, der Plattform der Abteilung „Innovative Informationssysteme und Publikationstechnologien“¹. Über sie werden sieben weitere Angebote für die Abfrage von Daten bereitgestellt:

¹<http://zbw.eu/labs/de/project/econ-ws>

- **suggest** - Der Endpunkt suggest liefert Vorschläge für Konzepte für die Eingabe von Begriffen, wobei sowohl bevorzugte Terme (prefLabel) als auch die alternativen Labels altLabel berücksichtigt werden. Zurückgegeben werden jeweils die URI der Konzepte, sowie die übereinstimmenden Terme und der jeweilige bevorzugte Begriff. In Abbildung 4.5 ist ein Konzept in einem Ausschnitt aus einer Antwort im JSON-Format abgebildet.

```
{
  "term": { "type": "literal" , "xml:lang": "en" , "value": "
    ↳ Acceptance of new products" } ,
  "concept": { "type": "uri" , "value": "http://zbw.eu/stw/
    ↳ descriptor/29817-6" } ,
  "prefLabel": { "type": "literal" , "xml:lang": "en" , "value
    ↳ ": "Innovation adoption" }
} ,
```

Abbildung 4.5: Rückgabe der „suggest“-Funktion des STW

- **suggest2** - Dieser Endpunkt dient der Suche nach Einträgen in den Personen- und Körperschaftsdatensätzen. Es muss ein dataset (econ_pers oder econ_corp) angegeben werden. Zurückgegeben werden Konzepte mit ihren bevorzugten Namen, der ihnen zugehörigen DNB-URIs und ihrem Eintrag im STW.

```
{
  "concept": { "type": "uri" , "value": "http://d-nb.info/
    ↳ gnd/10140798-1" } ,
  "prefLabel": { "type": "literal" , "value": "Deutsche
    ↳ Nationalbibliothek" } ,
  "prefName": { "type": "literal" , "value": "Deutsche
    ↳ Nationalbibliothek" }
}
```

Abbildung 4.6: Rückgabe der „suggest2“-Funktion des STW

- **synonyms** - Diese Option ermöglicht die Suche nach den Synonymen eines Begriffs sowohl auf Deutsch als auch auf Englisch.

```

{
  "synonym": { "type": "literal" , "xml:lang": "de" , "value":
    ↪ "brennstoffzelle" }
} ,
{
  "synonym": { "type": "literal" , "xml:lang": "en" , "value":
    ↪ "fuel cell" }
} ,
{
  "synonym": { "type": "literal" , "xml:lang": "de" , "value":
    ↪ "brennstoffzellentechnologie" }
}

```

Abbildung 4.7: JSON-Darstellung von Synonymen im STW

- **combined1** - Die Funktion `combined1` sucht für einen gegebenen Begriff über alle `prefLabel` und `hiddenLabel` nach Ergebnissen. Zusätzlich zu den gefundenen Konzepten werden die mit ihnen in Relation stehenden Konzepte ebenfalls zurückgegeben. In Abbildung 4.8 wird der Ausschnitt dargestellt, der das gefundene primäre Ergebniskonzept umfasst. Wie dort zu erkennen ist, wird in diesem auch über das Feld `related` auf die nicht im Ausschnitt enthaltenen verknüpften Konzepte verwiesen.

```

{
  "@id": "http://zbw.eu/stw/descriptor/18250-1",
  "@type": [
    "Found"
  ],
  "hiddenLabel": [
    {
      "@language": "de",
      "@value": "brennstoffzellentechnologie"
    },
    {
      "@language": "en",
      "@value": "fuel cell"
    },
    {
      "@language": "de",
      "@value": "brennstoffzelle"
    }
  ],
  "prefLabel": [
    {
      "@language": "en",
      "@value": "Fuel cell"
    },
    {
      "@language": "de",
      "@value": "Brennstoffzelle"
    }
  ],
  "related": [
    {
      "@id": "http://zbw.eu/stw/descriptor/14634-3"
    },
    {
      "@id": "http://zbw.eu/stw/descriptor/14173-6"
    }
  ]
}

```

Abbildung 4.8: JSON-Serialisierung eines SKOS-Concepts im STW

- **mappings** - Um die mit einem STW-Konzept verbundenen externen Ressourcen zu erkunden, wird der Endpunkt mappings bereitgestellt. Hiermit kann für eine gegebene URI eine Übersicht über alle mit dem Konzept verbundene Ressourcen erstellt werden (s. Abb. 4.9). Dies funktioniert sowohl für STW-Konzepte als auch für die externen Ressourcen wie etwa Einträge der Gemeinsamen Normdatei.

```

{
  "concept": { "type": "uri" , "value": "http://zbw.eu/stw/
    ↳ descriptor/18250-1" } ,
  "prefLabel": { "type": "literal" , "xml:lang": "de" , "value
    ↳ ": "Brennstoffzelle" } ,
  "relation": { "type": "uri" , "value": "http://www.w3.org
    ↳ /2004/02/skos/core#exactMatch" } ,
  "targetPrefLabel": { "type": "literal" , "xml:lang": "de" ,
    ↳ "value": "Brennstoffzelle" } ,
  "targetConcept": { "type": "uri" , "value": "http://d-nb.
    ↳ info/gnd/4008195-3" } ,
  "target": { "type": "uri" , "value": "http://zbw.eu/stw/
    ↳ mapping/gnd/target" }
}

```

Abbildung 4.9: Rückgabe der „mapping“-Funktion des STW

- **concept-history** - Suche nach Änderungen in einem Konzept per URI
(Es konnte kein Konzept mit relevanten Änderungen gefunden werden.)
- **assigned-descriptors** - Als Hilfs-Service bezeichnet, bietet dieser Endpunkt die Möglichkeit einer Suche nach Deskriptoren über die ISBN von verzeichneten Medien im Katalog der ZBW.

(Anfrage führten bei Untersuchung zu einem internen Fehler.)

Die Anzeige von Konzepten im STW ist in den RDF-Syntaxen XML/RDF und Turtle verfügbar. In Abb. 4.10 ist der Auszug eines Deskriptors aus einer Turtle-Antwort der STW-API dargestellt.

```

<descriptor/18250-1>
  gbv:gvkppn "091352878"^^xsd:string ;
  a skos:Concept, zbwest:Descriptor ;
  rdfs:isDefinedBy <descriptor/18250-1/about> ;
  skos:altLabel "Brennstoffzellentechnologie"@de ;
  skos:broader <descriptor/14473-1>, <descriptor/18353-5>, <thsys
    ↪ /70135>, <thsys/70538>, <thsys/71016> ;
  skos:exactMatch <http://d-nb.info/gnd/4008195-3>, <http://
    ↪ dbpedia.org/resource/Fuel_cell> ;
  skos:inScheme <../stw> ;
  skos:prefLabel "Brennstoffzelle"@de, "Fuel cell"@en ;
  skos:related <descriptor/14173-6>, <descriptor/14634-3> ;
  foaf:page <http://www.econbiz.de/subject/Fuel%2Bcell/>, <http://
    ↪ www.econbiz.de/thema/Brennstoffzelle/>, <../econis/search/
    ↪ descriptor/Brennstoffzelle>, <../econis/search/descriptor/
    ↪ Fuel%20cell> ;
  zbwest:indexedItem <http://www.econbiz.de/subject/Fuel%2Bcell/>,
    ↪ <http://www.econbiz.de/thema/Brennstoffzelle/>, <../
    ↪ econis/search/descriptor/Brennstoffzelle>, <../econis/
    ↪ search/descriptor/Fuel%20cell> .

```

Abbildung 4.10: Turtle-Repräsentation eines STW-Deskriptors

Für die Ergebnisse der Abfragen der Web Services sind weitere Formate verfügbar, unter anderem JSON-LD und N-Triples. Diese sind jedoch in der Dokumentation als experimentell gekennzeichnet (Deu16b).

4.3 Library of Congress Subject Headings

Die LCSH sind ein Thesaurus der Library of Congress mit etwa 340.000 Index-terminen (vgl. Lib16). Dieser wird von vielen Bibliotheken vor allem im englischen Sprachraum für die Sacherschließung verwendet. Die Subject Headings werden seit 1898 gepflegt und regelmäßig aktualisiert (vgl. Lib). Jedem Konzept der LCSH ist eine Library of Congress Control Number (LCCN) zugeordnet, dem Identifikator für alle Bestände der Library of Congress. Diese Nummer beginnt mit einem alphabetischen Prefix, gefolgt von einer vierstelligen Jahreszahl und einer sechsstelligen Ordnungszahl. Dabei findet der alphabetische Prefix nur bei Normdaten Anwendung und wird bei Freilassung durch Unterstriche ersetzt. Die

Library of Congress ist verantwortlich für die Entwicklung des Metadata Authority Description Schema (MADS), eines XML-Schemas für die bibliographische Erschließung von Normdaten, welches außerdem über eine RDF-Ontologie MADS/RDF verfügt. Das Schema wird in seiner Dokumentation als ein Datenmodell für Norm- und Vokabulardaten bezeichnet und verfügt über 58 Klassen und 93 Eigenschaften. Es wird ebenfalls hingewiesen auf sein Zusammenhang mit dem SKOS-Standard. MADS/RDF ist demnach eng mit SKOS verbunden, jedoch wird Letzteres für den Bedarf einer Normdatei für zu ungenau befunden. Es wird darauf hingewiesen, dass ein vollständiges Mapping von MADS/RDF auf SKOS vorhanden ist (vgl. [Lib15](#)). In Abbildung 4.11 ist die SKOS-Repräsentation eines Datensatzes dargestellt.

```
{
  "@id": "http://id.loc.gov/authorities/subjects/sh85052220",
  "@type": ["http://www.w3.org/2004/02/skos/core#Concept"],
  "http://www.w3.org/2004/02/skos/core#prefLabel": [{
    "@language": "en",
    "@value": "Fuel cells"
  }],
  "http://www.w3.org/2004/02/skos/core#broader": [{
    "@id": "http://id.loc.gov/authorities/subjects/sh85038254"
  }, {
    "@id": "http://id.loc.gov/authorities/subjects/sh85041588"
  }, {
    "@id": "http://id.loc.gov/authorities/subjects/sh85041916"
  }, {
    "@id": "http://id.loc.gov/authorities/subjects/sh85042105"
  }],
  "http://www.w3.org/2004/02/skos/core#narrower": [{
    "@id": "http://id.loc.gov/authorities/subjects/sh94002579"
  }, {
    "@id": "http://id.loc.gov/authorities/subjects/sh95004288"
  }, {
    "@id": "http://id.loc.gov/authorities/subjects/sh85052219"
  }, {
    "@id": "http://id.loc.gov/authorities/subjects/
    ↪ sh2008006260"
  }, {
    "@id": "http://id.loc.gov/authorities/subjects/
    ↪ sh200700667"
  }],
}
```



```

    "http://www.w3.org/2004/02/skos/core#exactMatch": [{
      "@id": "http://lod.nal.usda.gov/nalt/135578"
    }],
    "http://www.w3.org/2004/02/skos/core#closeMatch": [{
      "@id": "http://d-nb.info/gnd/4008195-3"
    }], {
      "@id": "http://data.bnf.fr/ark:/12148/cb11944532n"
    }],
    "http://www.w3.org/2004/02/skos/core#inScheme": [{
      "@id": "http://id.loc.gov/authorities/subjects"
    }],
    "http://www.w3.org/2004/02/skos/core#changeNote": [{
      "@id": "_:b32authoritiessubjectssh85052220"
    }], {
      "@id": "_:b40authoritiessubjectssh85052220"
    }
  ]
},

```

Abbildung 4.11: SKOS-Repräsentation eines LCSH-Konzepts

Für die Subject Headers wird lediglich die Möglichkeit angeboten, per Angabe ihrer LCCN ein Konzept abzufragen, jedoch bietet die Schnittstelle viele Optionen bezüglich der zur Verfügung stehenden Formate an. So stehen für jedes Konzept elf Rückgabeformate zur Verfügung, die aus Verschiedenen Kombinationen von verwendeten Ontologien und Formaten bestehen.

4.4 Gemeinsame Normdatei

Die GND ist die von der Deutschen Nationalbibliothek geführte Normdatei für Personen, Körperschaften, Kongresse, Geografika, Schlagwörter und Werktitel. Sie bildet den Standard für die formale Erfassung von Dokumenten durch Bibliotheken für den gesamten deutschsprachigen Raum und umfasst zu Zeit ca. 11,3 Mio. Datensätze (vgl. [Deu16a](#)). Die GND ist aus einer 2012 erfolgten Zusammenlegung der zuvor getrennt geführten Personennamendatei (PND), Gemeinsamen Körperschaftsdatei (GKD), Schlagwortnormdatei (SWD) und der Einheitssachtitel-Datei des Deutschen Musikarchivs (DMA-EST-Datei) hervorgegangen (vgl. [Nig12](#), S.43). Die Normdatei fungiert als Standard für die Erschließung von Medien im deutschen Sprachraum und wird kooperativ durch die DNB, die deutschen

Bibliotheksverbünde und einzelne Bibliotheken gepflegt. Für die Darstellung von Normdaten verwendet die DNB hauptsächlich ihre eigene "GND Ontology". Dieser RDF-Namespaces besteht zur Zeit aus 228 Eigenschaften und 57 Klassen (vgl. Haf16), die den Ausdrucksbedarf für die unterschiedlichen Typen von Normdaten im Datenbestand der GND decken soll.

Die vom Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) betriebene Plattform „lobid“ (kurz für „Linking Open Bibliographic Data“) stellt eine API für die GND als auch für Titel-, Inventar- und Organisationsdaten der mit dem hbz verbundenen Institutionen zur Verfügung². Die Schnittstelle verfügt über sechs Endpunkte, für die jeweils vier verschiedene Suchparameter spezifiziert werden können. Die auf GND-Daten beschränkten Endpunkte sind hierbei „person“ und „subject“. Der erste dieser Endpunkte verweist nur auf Personendaten, während unter „subject“ neben Schlagwörtern auch die Körperschaftsdatei suchbar ist. Drei der Parameter bieten Beschränkungen auf die Suche per GND-ID (id) oder Termen (name), sowie einer Suche über alle Felder der Datensätze (q). Der vierte Parameter format bestimmt die Art der Rückgabe. Der Wert short bewirkt die Rückgabe einer Liste von bevorzugten Begriffen, ids gibt diese zusammen mit den jeweiligen URIs der Konzepte zurück. Die vollständige Anzeige eines GND-Konzeptes wird durch den Wert full zur Verfügung gestellt. Der Wert negotiate dient der Rückgabe in anderen Formaten, welche durch die Spezifizierung des Access-Headers gewählt werden können. Das standardmäßig zurückgelieferte Format ist JSON-LD (s. Abb. 4.12).

```
{
  "primaryTopic" : "http://d-nb.info/gnd/4008195-3",
  "@graph" : [{
    "http://www.w3.org/2004/02/skos/core#exactMatch" : {
      "@id" : "http://zbw.eu/stw/descriptor/18250-1"
    },
    "@type" : [
      "http://d-nb.info/standards/elementset/gnd#AuthorityResource",
      "http://d-nb.info/standards/elementset/gnd#SubjectHeading"
    ]
  }
],
```

²Die Dokumentation der Schnittstelle ist unter <http://lobid.org/api> erreichbar.

```

"gnSubjectCategory" : [
  "http://d-nb.info/standards/vocab/gnd/gnd-sc#31.9a",
  "http://d-nb.info/standards/vocab/gnd/gnd-sc#31.5"
],
"oldAuthorityNumber" : "(DE-588c)4008195-3",
"broaderTermGeneral" : [
  "http://d-nb.info/gnd/4014720-4",
  "http://d-nb.info/gnd/4155907-1"
],
"preferredNameForTheSubjectHeading" : "Brennstoffzelle",
"relatedDdcWithDegreeOfDeterminacy4" : "http://d-nb.info/ddc/class
  ↪ /621.312429",
"@id" : "http://d-nb.info/gnd/4008195-3",
"gnIdentifier" : "4008195-3",
"preferredName" : "Brennstoffzelle",
"variantName" : [
  "Elektrochemische Brennstoffzelle",
  "Brennstoffelement <Direktumwandlung>",
  "Brennstoffelement <Energiedirektumwandlung>"
],
"variantNameForTheSubjectHeading" : [
  "Brennstoffelement <Direktumwandlung>",
  "Brennstoffelement <Energiedirektumwandlung>",
  "Elektrochemische Brennstoffzelle"
]
}],
"@id" : "http://d-nb.info/gnd/4008195-3/about",
"@context" : "http://lobid.org/context/gnd.json"
}

```

Abbildung 4.12: GND-Eintrag in seiner JSON-Serialisierung in lobid

In diesem Beispiel sind sehr gut einige der besonders komplexen und spezifischen Eigenschaften der Ontologie zu erkennen, etwa der Stärke der Relation zu einem spezifischen DDC-Term (`relatedDdcWithDegreeOfDeterminacy4`) oder dem Verweis auf die ID des Konzepts in der früheren Schlagwortnormdatei (`oldAuthorityNumber`). Auch ersichtlich ist jedoch, dass sich einige der Eigenschaften auch leicht durch das schlankere SKOS-Format ausdrücken lassen könnten. Dazu gehören etwa `variantName`, `preferredName` und `broaderTermGeneral`.

5 Vergleich der Schnittstellen

Die in Kapitel 4 untersuchten Schnittstellen sind allesamt Angebote für das Retrieval von kontrollierten Vokabularen von Informationseinrichtungen. Sie ähneln sich in ihrer Umsetzung in bestimmten Aspekten, jedoch können auch deutliche Unterschiede in den Konzeptionen und der vorgesehenen Nutzungsrahmen der einzelnen APIs festgestellt werden. In diesem Kapitel werden die Eigenschaften der Angebote nach einigen Gesichtspunkten miteinander verglichen und mit Hinblick auf deren Einfluss auf die Interoperabilität solcher Systeme analysiert. Zunächst werden hierzu die generellen Nutzungsoptionen der Schnittstellen anhand ihrer Abfrageparameter untersucht und die in Tabelle 5.1 verzeichneten Onlinedokumentationen verglichen. Der nächste Punkt von Interesse sind die Rückgabeformate der einzelnen APIs, und damit verbunden auch die verwendeten RDF-Ontologien. Die letzten beiden Kriterien beziehen sich dann auf den Umgang mit internen und externen Relationen, die einen grundlegenden Bestandteil solcher Informationssysteme darstellen.

Schnittstelle	Dokumentation
RVK	http://rvk.uni-regensburg.de/api/
lobid (GND)	http://lobid.org/api
LCSH	-
STW	http://zbw.eu/beta/econ-ws/about

Tabelle 5.1: Dokumentation der APIs

5.1 Abfrageparameter und Dokumentation

Die Auswahl von Abfrageparameter der verschiedenen APIs ist von Interesse, um die beabsichtigten Nutzungsszenarien und Entwicklungsphasen der Schnittstellen einschätzen zu können. Hier zeigen sich für die Untersuchungsobjekte zunächst Unterschiede, was die erwarteten Anwendungsmuster betrifft. Das einfachste Beispiel der ausgewählten Systeme ist die API der Library of Congress. Diese stellt lediglich eine vollständige Repräsentation eines einzelnen Konzepts durch die Angabe der LCCN zur Verfügung. Damit ist es zum Beispiel nicht möglich, eine maschinelle Begriffssuche durchzuführen. Für diese Schnittstelle gibt es keine weitere Dokumentation, vorhanden ist lediglich eine Auswahl der möglichen

Rückgabeformate in der HTML-Ansicht eines Konzeptes. Dies ist zwar der gleiche Ansatz wie bei der Konzeptanzeige des STW, jedoch bietet die ZBW noch weitere Retrievaloptionen über ihre Web Services an. Diese Werkzeuge und die Vermerke in der Dokumentation der STW-API bezüglich des experimentellen Status einiger Methoden lassen darauf schließen, dass die Entwicklungsbemühungen der ZBW in diesem Bereich deutlich aktiver sind. Im Gegensatz dazu steht die RVK-API, von deren dokumentierten Beispielen einige nicht funktionieren und deren neue Version der API (noch) gar nicht dokumentiert ist.

Die beiden Endpunkte der Plattform lobid des hbz für die Abfrage von GND-Daten dienen der feineren Untergliederung der Normdatei. Diese hat mit ihren 11,3 Mio. Datensätzen im Vergleich zu den restlichen untersuchten Vokabularen einen deutlich größeren Umfang, was vor allem an der Verzeichnung von Personen- und Körperschaftsdaten liegt. Die Schnittstelle ist unter den verglichenen KOS die Einzige, welche für die Spezifizierung des Rückgabeformats lediglich die Möglichkeit der Festlegung eines HTML Access-Headers gibt. In diesem Aspekt verfahren jedoch auch alle anderen Schnittstellen unterschiedlich. Während die RVK-API das Format als Endpunkt definiert, wird die Festlegung im Fall der Web Services der ZBW durch den Parameter `output` durchgeführt. Beim Abruf der Konzepte des STW wiederum, die über eine getrennte API verfügen, wird dahingegen an die URI des Konzeptes lediglich die entsprechende Dateiendung angehängt. Ähnlich funktioniert auch die Schnittstelle der LCSH, wobei hier auf Grund der verschiedenen Arten gleicher Formate teilweise ein weiteres Kürzel vorangestellt wird. Ein Beispiel hierfür sind die Rückgabeformate „JSON (MADS/RDF and SKOS/RDF)“ und „SKOS - JSON“, welche die Endungen `.json` bzw. `.skos.json` haben.

5.2 Nutzung von Rückgabeformaten

In der Tabelle 5.2 sind die verfügbaren Rückgabeformate der APIs zusammengefasst. Aus der Übersicht lässt sich erkennen, dass alle der Untersuchungsobjekte mindestens von den beiden vorherrschenden Webformaten JSON und XML Gebrauch machen, jedoch lässt sich schon hier die Abgrenzung der RVK-API zu den restlichen Schnittstellen erkennen. Mit Ausnahme der RVK-Schnittstelle basieren alle Rückgaben auf einer RDF-Datenbasis, dementsprechend werden in diesen Fällen auch überwiegend RDF-Syntaxen wiedergegeben.

Schnittstelle	Rückgabeformate
RVK	JSON, XML
lobid (GND)	JSON-LD, RDF-XML, N-Triples, Turtle, N3
LCSH	JSON-LD, RDF-XML, N-Triples
STW	RDF-XML, Turtle, JSON, N-Triples

Tabelle 5.2: Verfügbare Rückgabeformate der Schnittstellen

Für die genauere Untersuchung der Umsetzungen der einzelnen Schnittstellen wurden Beispielabfragen getätigt, um verschiedene Herangehensweisen der Anbieter miteinander vergleichen zu können. Die Ergebnisse beschränken sich im Folgenden auf die JSON-Serialisierungen der APIs, wobei angemerkt werden muss, dass sich diese Eigenschaften wegen der zugrunde liegenden Datenmodelle im Allgemeinen auf die restlichen Formate übertragen lassen.

Die Serialisierungen der Schnittstellen lassen sich zunächst in die Verwendung von traditionellem JSON und der Repräsentation von RDF-Daten durch JSON-LD aufteilen. Ersteres wird sowohl durch die RVK-API als auch einigen Rückgabeformaten der Web Services der ZBW praktiziert. Während die RVK ihre eigenen Feldbenennungen verwendet (Abb. 5.1), bildet – wie in Abbildung 5.2 zu erkennen – SKOS die Grundlage für einen Teil der Felder im STW. Dies kann darauf zurückgeführt werden, dass diese Daten aus einem RDF-Datensatz generiert wurden.

```

{
  "host": "rvk.uni-regensburg.de",
  "request": "/Portal_API/json/node/VN%206050%20-%20VN%206059",
  "node": {
    "notation": "VN 6050 - VN 6059",
    "benennung": "Brennstoffzellen; Brennstoffe Batterien; Angewandte
      ↳ Elektrochemie; Solarzellen; Wasserstoff als Energieträger",
    "bemerkung": "s.a. VE 6300 Elektrochemie; s.a. VE 6350 Elektrolyte; s.
      ↳ a. VG 8000 Elektroanalyse; s.a. ZP 4120",
    "verweis": "",
    "has_children": "yes",
    "register": ["Batterie", "Brennstoff", "Brennstoffzelle", "
      ↳ Elektrochemie", "Energieträger", "Solarzelle", "Technik", "
      ↳ Wasserstoff"]
  },
  "root_name": "UB Regensburg"
}

```

Abbildung 5.1: JSON-Serialisierung eines Konzepts der RVK-API

```

"head": {
  "vars": [ "concept" , "prefLabel" , "relation" , "targetPrefLabel" , "
    ↳ targetConcept" , "target" ]
} ,
"results": {
  "bindings": [
    {
      "concept": { "type": "uri" , "value": "http://d-nb.info/gnd
        ↳ /4291458-9" } ,
      "prefLabel": { "type": "literal" , "xml:lang": "de" , "value": "
        ↳ Grenzproduktivität" } ,
      "relation": { "type": "uri" , "value": "http://www.w3.org/2004/02/
        ↳ skos/core#exactMatch" } ,
      "targetPrefLabel": { "type": "literal" , "xml:lang": "de" , "value
        ↳ ": "Grenzproduktivität" } ,
      "targetConcept": { "type": "uri" , "value": "http://zbw.eu/stw/
        ↳ descriptor/29921-4" } ,
      "target": { "type": "uri" , "value": "http://zbw.eu/stw" }
    }
  ]
}

```

Abbildung 5.2: JSON-Serialisierung eines Mappings im STW

Die Serialisierungen von RDF in JSON-LD bieten ebenfalls interessante Unterschiede. Alle drei untersuchten Rückgaben zeigen einen unterschiedlichen Umgang mit RDF-Vokabularen. Die Rückgaben im SKOS-Format der LCSH benutzen die ungekürzten URIs der Klassen und Eigenschaften der jeweiligen Ontologien als Feldname und vergeben dazu jeweils durch die JSON-LD-Präfixe @value oder @id die Subjekte der Statements:

```
"http://www.w3.org/2004/02/skos/core#prefLabel": [{
  "@language": "en",
  "@value": "Fuel cells"
}],
```

Dementgegen verwenden die Web Services des STW das Keyword @context, um die einzelnen in dem Objekt verwendeten Felder zu identifizieren:

```
"@context": {
  "prefLabel": "http://www.w3.org/2004/02/skos/core#prefLabel",
  "exactMatch": "http://www.w3.org/2004/02/skos/core#exactMatch",
  "closeMatch": "http://www.w3.org/2004/02/skos/core#closeMatch",
  "relatedMatch": "http://www.w3.org/2004/02/skos/core#
    ↳ relatedMatch",
  "narrowMatch": "http://www.w3.org/2004/02/skos/core#narrowMatch
    ↳ ",
  "broadMatch": "http://www.w3.org/2004/02/skos/core#broadMatch",
  "isPartOf": "http://purl.org/dc/terms/isPartOf"
}
```

Das Format der GND-Plattform lobid.org macht ebenfalls Gebrauch von @context, verweist hier jedoch lediglich auf eine externe JSON-Datei, in der die verwendete GND Ontology beschrieben wird:

```
"@context" : "http://lobid.org/context/gnd.json"
```

Bezüglich des Aufbaus der RDF-Graphen und der Verwendung des Feldes @graph stimmen die Formate der STW und der GND miteinander überein in ihrer Nutzung des Präfix. Für die JSON-Formate der LCSH wird dies weggelassen und eine flache Struktur mit RDF-Statements zurückgegeben.

5.3 Verwendete Ontologien

Da außer der RVK-API alle Schnittstellen auf RDF basieren, ist die Betrachtung der verwendeten Ontologien interessant, um einen Eindruck über die Ähnlichkeiten der Normdatenschnittstellen zu erlangen. In Tabelle 5.3 sind alle von den Schnittstellen verwendeten RDF-Schemaformate aufgelistet. Zu erkennen sind grundsätzliche Überschneidungen, aber auch systemspezifische Ontologien, die im weiteren behandelt werden. SKOS wird beispielsweise als offizieller Standard des W3C von allen Institutionen in unterschiedlichem Umfang verwendet.

Schnittstelle	Verwendete Ontologien
RVK	-
LCSH	SKOS, Changeset, MADSRDF,
STW	SKOS, dcterms, foaf, zbwtext, gbv
lobid (GND)	gndo, foaf, SKOS

Tabelle 5.3: Verwendete RDF-Ontologien der Schnittstellen

Von den drei in RDF-Metadaten vorhandenen Systemen ähneln sich die Rückgaben des ZBW und das SKOS-Format der Library of Congress am meisten. Beide Formate benutzen überwiegend SKOS für die strukturelle Darstellung der Datensätze. Beide Datenmodelle werden zunächst als ein `skos:Concept` definiert und der bevorzugte Term wird als `skos:prefLabel` mit Sprachcodes dargestellt. Ebenso werden die SKOS-Terme für semantische Relationen – `broader`, `narrower` und `related` – sowie die „Mapping Properties“ von SKOS von beiden Systemen verwendet. Die Daten der Subject Headings weisen außerdem das Feld `skos:ChangeNote` für die Verzeichnung von Änderungen an einem Konzept auf. In dieser Verbindung wird auch auf die Ontologie „Changeset“ zurückgegriffen, die unter anderem Felder für Autoren, Daten und Notizen von Änderungen bereitstellt. Die STW nutzt in Ergänzung einen eigenen Namespace `zbwext`¹, mit dem unter anderem auf Stellen in dem Katalog der ZBW durch das Feld `zbwext:indexedTerm` verwiesen wird. Wie in Abbildung 5.3 zu erkennen ist, wird das Feld parallel zu dem Tag `foaf:page` der Foaf-Ontologie verwendet. Zusätzlich vergibt die ZBW für das

¹Die Spezifikation der „zbw-extensions“ ist unter <http://zbw.eu/namespaces/zbw-extensions/zbw-extensions.rdf> zu erreichen

Konzept eine Pica Produktionsnummer (PPN) unter der Verwendung des Felds `gbv:gvpkn` der GBV-Ontologie².

```
foaf:page <http://www.econbiz.de/subject/Fuel%2Bcell/>, <http://  
  ↳ www.econbiz.de/thema/Brennstoffzelle/>, <../econis/search/  
  ↳ descriptor/Brennstoffzelle>, <../econis/search/descriptor/  
  ↳ Fuel%20cell> ;  
zbwext:indexedItem <http://www.econbiz.de/subject/Fuel%2Bcell/>,  
  ↳ <http://www.econbiz.de/thema/Brennstoffzelle/>, <../  
  ↳ econis/search/descriptor/Brennstoffzelle>, <../econis/  
  ↳ search/descriptor/Fuel%20cell> .
```

Abbildung 5.3: Verwendung der `zbwext`-Ontologie zum Verweis auf externen Gebrauch

Im Gegensatz zu dieser Verwendung von SKOS als bevorzugte Ontologie stehen die anderen Formate der LCSH und GND. Die Library of Congress bietet neben der Darstellung des Subject Headings als SKOS-Konzept auch die Präsentation in dem von der Institution selbst entworfenen MADSRDF-Format an. Dies ist zwar auch vollständig auf SKOS gemappt, bietet jedoch deutlich mehr Felder für die Beschreibung von Personen- und Körperschaftsdaten, sowie für Geografika. Beispiele für solche Felder sind zum Beispiel Geburtsdaten (`madsrdf:birthDate` und `madsrdf:birthPlace`) oder Anschriften (`madsrdf:streetAddress`). Der gleiche Aspekt trifft auch auf die im Zusammenhang mit der GND verwendete GND Ontology zu. Der Rahmen dieses Namespaces ist im Bezug auf die vorhandenen Eigenschaften sogar noch größer (228 GND-Eigenschaften im Vergleich zu 93 MADS-Eigenschaften). Im Gegensatz zu MADS/RDF besteht auch kein offizielles Mapping der gesamten GND Ontology auf SKOS, obwohl es hier auch schon Experimente zu gab, etwa das Mapping-Projekt von Joachim Neubert der ZBW (Neu16). In der RDF-Spezifikation der Ontologie³ finden sich allerdings Äquivalenzverknüpfungen über die `owl:equivalentClass`, mit denen die SKOS-Felder `Concept`, `prefLabel`, `altLabel`, `narrower` und `broader` besetzt werden. Des Weiteren gibt es die so genannten „GND Subject Categories“, eine Liste von zur Zeit 845 Begriffen die von der DNB als `skos:Concepts` definiert wurden (Deu). Die Anpassung von Konzepten an die Strukturen anderer Ontologien für

²Dokumentiert unter <http://gbv.github.io/gbvontology/gbvontology.html>

³<http://d-nb.info/standards/elementset/gnd>

die Normdatei wird weiterhin als Ziel der DNB beschrieben, das im Rahmen der teilhabenden Institutionen vorangetrieben wird (Haf12). Zur Zeit sind die einzigen in der Kontextspezifikation der lobid-API⁴ definierten Felder jedoch die Friend of a Friend (FoaF)-Eigenschaften `foaf:page` und `primaryTopic`. Die einzige Verwendung von SKOS ist die Verknüpfung mit dem STW über die Mapping Properties `skos:closeMatch` und `skos:exactMatch`.

5.4 Interne Relationen

Die internen Relationen der Vokabulare werden durch die Schnittstellen in unterschiedlicher Weise behandelt. Die größte Abweichung von allen anderen Angeboten bietet hierbei die RVK-API. Die Auslassung jeglicher Verlinkung der vorhandenen hierarchischen Beziehungen zeugen davon, dass hier bei der Entwicklung der Schnittstelle die Maxime des Semantic Web nicht wirklich in Erwägung gezogen wurden. Einzig die Abfrage nach den `ancestors` bietet einen Mehrwert im Kontext der monohierarchisch strukturierten Klassifikation. Durch die Anzeige aller übergeordeter Klassen lässt sich derer Ursprung mit diesem Mittel gut verfolgen. Das Bedürfnis nach solch einer Funktion lässt sich durch die Notationsbildung der RVK erklären, da zum Beispiel im Gegensatz zu dezimal aufgebauten Systematiken die direkte Abfolge von Oberklassen nicht aus der Notation erkennen lässt.

Einander sehr ähnlich im Umgang mit hierarchischen und assoziativen Relationen sind der STW und die SKOS-Darstellung der LCSH. Beide benutzen die Eigenschaften `skos:broader`, `skos:narrower` und `skos:related`, um über URIs auf andere Konzepte zu verweisen. Das Standardformat der Library of Congress (LoC) liefert SKOS parallel zu dem eigenen MADS-Format zurück. Zu erkennen ist hier die Äquivalenz der Eigenschaften der beiden Ontologien. Dabei werden in der `madsrdf`-Spezifikation die Eigenschaften `madsrdf:hasBroaderAuthority`, `madsrdf:hasNarrowerAuthority` und `madsrdf:hasRelatedAuthority` verwendet. Letztere ist im Gegensatz zum STW auf die SKOS-Property `skos:semanticRelation` gemappt.

⁴<http://lobid.org/context/gnd.json>

Für die GND-Schnittstelle wird in diesem Rahmen die entsprechende Eigenschaft `broaderTermGeneral` analog verwendet, jedoch verfügt die GND Ontology über mehrere (nicht auf SKOS gemappte) Assoziationsrelationen wie etwa `relatedConferenceOrEvent` oder `relatedFamily`. Das dem SKOS-Element `skos:narrower` entsprechende `narrowerTermGeneral` der GND Ontology wird zudem in der Kontextdatei⁵ der lobid-API nicht einmal erwähnt.

5.5 Verweise auf externe Ressourcen

Die Vernetzung von Ressourcen unterschiedlicher Herkunft ist ein integraler Bestandteil der Idee des Semantic Web (BHBL09). Im Bezug auf kontrollierte Vokabulare von Informationseinrichtungen sind hierbei zunächst Verbindungen mit anderen einflussreichen KOS logisch. In der Auswahl der Schnittstellen sticht unter diesem Gesichtspunkt, wie auch unter einigen anderen Aspekten, zunächst die RVK-API heraus. In dem Datenbestand der RVK lassen sich keine Anzeichen für eine eventuelle Vernetzung mit anderen Anbietern von Wissensorganisationssystemen erkennen. Alle anderen Untersuchungsobjekte verfügen hingegen zumindest über einzelne externe Beziehungen. Als bedeutende Vokabulare im deutschsprachigen Raum ist es nicht verwunderlich, dass die STW und die GND gegenseitig aufeinander verweisen. Der Wirtschaftsthesaurus bietet zusätzlich eine Verknüpfung zur DBPedia Ontology an, mit der Wikipedia-Artikel verbunden sind. Durch lobid wird auf die institutionseigenen GND Subject Categories verwiesen:

```
"gndSubjectCategory" : [  
  "http://d-nb.info/standards/vocab/gnd/gnd-sc#31.9a",  
  "http://d-nb.info/standards/vocab/gnd/gnd-sc#31.5"  
],
```

Im Rahmen des CrissCross-Projekts der DNB entstanden 2006 und 2010 Mappings zwischen der GND (damals noch SWD) und der deutschen Version der DDC, sowie zu den LCSH und der französischen Normdatei Répertoire d'autorité-matière encyclopédique et alphabétique unifié (Bee06). Während die Beziehungen zur DDC ebenfalls in den Daten der lobid-Schnittstelle vorhanden sind, fehlen die

⁵<http://lobid.org/context/gnd.json>

beiden anderen Verknüpfungen. Da diese generell existieren, lässt sich durch die Recherche auf dem Internetportal der DNB herausfinden⁶. Andersherum ist die Verknüpfung der beiden Vokabulare jedoch existent – die LCSH liefert Verknüpfungen sowohl zur GND als auch zu dem entsprechenden Datensatz der französischen Nationalbibliothek:

```
"http://www.w3.org/2004/02/skos/core#closeMatch": [{
  "@id": "http://d-nb.info/gnd/4008195-3"
}, {
  "@id": "http://data.bnf.fr/ark:/12148/cb11944532n"
}],
```

Zusammenfassend lässt sich festhalten, dass die untersuchten Institutionen (mit Ausnahme der RVK) einen guten Anfang bezüglich der Verknüpfung ihrer Systeme mit anderen Datenanbietern gemacht haben. Die Verknüpfung des STW mit der DBPedia und damit den Wikipedia-Seiten zu den jeweiligen Konzepten kann für die anderen Vokabulare als Vorbild gesehen werden. Die GND verfügt mit der bestehenden Konkordanz zu der bedeutenden amerikanischen Klassifikation DDC ein wichtiges Mittel, um die Internationalisierung von solchen Systemen voranzutreiben. Allerdings zeigt die Dauer des Projekts zur Erstellung dieser Verknüpfungen auch, was für einen Aufwand die Schaffung solcher Verbesserungen für die Institutionen darstellen können.

6 Fazit

Die durchgeführte Untersuchung der Webschnittstellen zeigt auf, dass die verschiedenen Institutionen unterschiedliche Wege bezüglich der Nutzbarmachung ihrer kontrollierten Vokabulare gegangen sind. Außerdem variieren diese im Bezug auf die (korrekte) Verwendung von Webstandards. Am Beispiel der RVK-API lässt sich eine Implementierung ohne Berücksichtigung der vorhandenen Standards bezüglich des Austauschs von Daten über das Web darstellen. Die Schwächen im Hinblick auf die Interoperabilität der API lassen sich sowohl bezüglich der Wahl der Datenstrukturen als auch den Mangel an Verknüpfungen zu externen Ressourcen erkennen. Die restlichen Schnittstellen genügen in deutlich größerem

⁶vgl. bspw. <http://d-nb.info/gnd/4008195-3> und <http://lobid.org/subject?id=4008195-3>

Ausmaß den Empfehlungen des W3C, obwohl auch hier einige Besonderheiten und Mängel aufgezeigt wurden. Auch zeigen die APIs der Gemeinsamen Normdatei und der Library of Congress Subject Headings, dass sich die Übertragung von solch komplexen Systemen in allgemeine Standards wie z.B. SKOS als sehr schwierig erweist. Dennoch können hier auch noch Verbesserungen bezüglich der bisherigen Umsetzungen vorgenommen werden. In diesem Zusammenhang soll auch der Bezug dieser Arbeit zu der Spezifizierung einer einheitlichen API für kontrollierte Vokabulare auf Basis einer JSON-LD-Syntax für SKOS erwähnt werden, die im Rahmen des Projekts „coli-conc“¹ der Verbundzentrale Göttingen entwickelt wird.

An diesem Punkt muss auch erwähnt werden, dass sich die Suche nach offenen APIs teilweise als schwierig herausgestellt hat. Während einige Anbieter viel genutzter Vokabulare (z.B. der DDC) lediglich geschlossene bzw. kostenpflichtige APIs anbieten, sind von kleineren Anbietern häufig gar keine Schnittstellen vorhanden. Hinzu kommt in einigen Fällen eine schlechte Sichtbarkeit der Angebote, etwa durch fehlende Verlinkung von der Homepage einer betreibenden Institution. Hierfür könnte auch der Ausbau von Verzeichnissen von Vokabularen und deren Schnittstellen eine große Hilfe darstellen. Ein hoffnungsvoller Kandidat dafür ist das Basel Register of Thesauri, Ontologies & Classifications (BARTOC)², über welches bereits eine Menge an Informationen bezüglich kontrollierter Vokabulare aller Art erhältlich sind.

Abschließend kann gesagt werden, dass in den Bereichen der Vernetzung und Bereitstellung von kontrollierten Vokabularen Fortschritte gemacht werden. Der Vorteil von APIs für den maschinellen Austausch solcher Daten wird von vielen Institutionen des Informationswesens anerkannt und an weiteren Verbesserungen der bestehenden Werkzeuge wird gearbeitet. Für die Unterstützung dieses Prozesses ist die Weiterentwicklung und Anerkennung von Standards im Bereich der digitalen Verwaltung von Bibliotheksmetadaten von großer Bedeutung.

¹<https://coli-conc.gbv.de/>

²<http://bartoc.org/>

Literaturverzeichnis

- [Bec14] Beckett, David: *RDF 1.1 N-Triples: A line-based syntax for an RDF graph*. <https://www.w3.org/TR/n-triples/>. Version: 25.02.2014. – zuletzt geprüft am: 09.08.2016
- [Bee06] Bee, Guido: CrissCross. In: *Dialog mit Bibliotheken* 18 (2006), Nr. 2, S. 23–24
- [Ber05] Bertram, Jutta: *Content and communication*. Bd. 2: *Einführung in die inhaltliche Erschließung: Grundlagen, Methoden, Instrumente*. Würzburg : Ergon-Verl., 2005. – ISBN 978–3899134421
- [BHBL09] Bizer, Christian ; Heath, Tom ; Berners-Lee, Tim: *Linked Data - The Story So Far*. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Version: 2009. – zuletzt geprüft am: 02.08.2016
- [Deu] Deutsche Nationalbibliothek: *GND Subject Categories*. <http://d-nb.info/standards/vocab/gnd/gnd-sc.html>. – zuletzt geprüft am: 09.08.2016
- [Deu15] Deutsche Zentralbibliothek für Wirtschaftswissenschaften: *Standard-Thesaurus Wirtschaft*. <http://zbw.eu/stw/version/latest/about>. Version: 15.06.2015. – zuletzt geprüft am: 12.04.2016
- [Deu16a] Deutsche Nationalbibliothek: *Jahresbericht 2015*. <http://d-nb.info/1102219754/34>. Version: 2016. – zuletzt geprüft am: 12.07.2016
- [Deu16b] Deutsche Zentralbibliothek für Wirtschaftswissenschaften: *Web Services for Economics*. <http://zbw.eu/beta/econ-ws/about>. Version: 03.05.2016. – zuletzt geprüft am: 05.08.2016
- [Fie00] Fieldings, Roy T.: *Architectural Styles and the Design of Network-based Software Architectures*. Irvine, CA, University of California, Diss.,

2000. https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf

- [Haf12] Haffner, Alexander: *Internationalisierung der GND durch das Semantic Web*. <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Berichte/internationalisierungDerGndDurchDasSemanticWeb.pdf>.
Version: 2012. – zuletzt geprüft am: 09.08.2016
- [Haf16] Haffner, Alexander: *GND Ontology*. <http://d-nb.info/standards/elementset/gnd>. Version: 12.01.2016. – zuletzt geprüft am: 12.01.2016
- [HKPR12] Hitzler, Pascal ; Krötzsch, Markus ; Parsia, Bijan ; Rudolph, Sebastian: *OWL 2 Web Ontology Language Primer (Second Edition)*. <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>. Version: 11.12.2012. – zuletzt geprüft am: 10.08.2016
- [HWN] Häusler, Ines ; Werr ; Naoka: *Die Regensburger Verbundklassifikation (RVK): Zur Dynamik einer Klassifikation*. http://www.vdb-online.org/veranstaltungen/543/5_werr.pdf. – zuletzt geprüft am: 10.08.2016
- [IS09] Isaac, Antoine ; Summers, Ed: *SKOS Simple Knowledge Organization System Primer*. <https://www.w3.org/TR/skos-primer/>. Version: 18.08.2009. – zuletzt geprüft am: 08.08.2016
- [KSS⁺14] Kuhlen, Rainer (Hrsg.) ; Semar, Wolfgang (Hrsg.) ; Strauch, Dietmar (Hrsg.) ; Laisiepen, Klaus (Hrsg.) ; Lutterbeck, Ernst (Hrsg.) ; Meyer-Uhlenried, Karl-Heinrich (Hrsg.): *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. 6., völlig neu gefasste Ausg. Berlin : de Gruyter, 2014. – ISBN 978-3110258257
- [Lib] Library of Congress: *Library of Congress Subject Headings*. <http://id.loc.gov/authorities/subjects>

- [Lib15] Library of Congress: *MADS/RDF Primer*. <http://www.loc.gov/standards/mads/rdf/>. Version: 28.10.2015. – zuletzt geprüft am: 09.08.2016
- [Lib16] Library of Congress: *Introduction*. <https://www.loc.gov/aba/publications/FreeLCSH/lcshintro.pdf>. Version: 2016. – zuletzt geprüft am: 12.07.2016
- [Neu16] Neubert, Joachim: *Turning the GND subject headings into a SKOS thesaurus: an experiment*. <http://zbw.eu/labs/de/blog/turning-the-gnd-subject-headings-into-a-skos-thesaurus-an-experiment>. Version: 30.03.2016. – zuletzt geprüft am: 09.08.2016
- [Nig12] Niggemann, Elisabeth (Hrsg.): *Hundert. Bd. 4: Leseraum: von Schmöckern bis Studieren: wie Worte Welten eröffnen*. Frankfurt am Main : Dt. Nationalbibliothek, 2012. – ISBN 978-3-941113-35-0
- [SLK⁺14] Sporny, Manu ; Longley, Dave ; Kellogg ; Gregg ; Lanthaler, Markus ; Lindström, Niklas: *JSON-LD 1.0: A JSON-based Serialization for Linked Data*. <https://www.w3.org/TR/json-ld>. Version: 25.02.2014. – zuletzt geprüft am: 10.08.2016
- [Spra] Spree, Ulrike: *Kontrollierte Vokabulare - Grundlagen*. <http://www2.bui.haw-hamburg.de/pers/ulrike.spree/remind/vokabulare.htm>. – zuletzt geprüft am: 12.04.2016
- [Sprb] Spree, Ulrike: *Thesaurus*. <http://www2.bui.haw-hamburg.de/pers/ulrike.spree/remind/thesaurus.htm>. – zuletzt geprüft am: 09.08.2016
- [Uni] Universität Regensburg: *Es begann vor 50 Jahren: Werdegang der RVK*. <http://rvk.uni-regensburg.de/44-ueber-die-rvk-regensburger-verbundklassifikation/geschichte-rvk-regensburger-verbundklassifikation-historie-entwicklung/57-werdegang-der-rvk>. – zuletzt geprüft am: 12.04.2016

[W3C14] W3C: *RDF 1.1 Primer*. <https://www.w3.org/TR/rdf11-primer/>.
Version: 24.06.2014. – zuletzt geprüft am: 06.08.2016

[Wik15] Wikipedia: *Regensburger Verbundklassifikation* — *Wikipedia, Die freie Enzyklopädie*. https://de.wikipedia.org/w/index.php?title=Regensburger_Verbundklassifikation&oldid=147081579.
Version: 2015

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die eingereichte Bachelor-/Masterarbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ort, Datum

Unterschrift