

Überführung landwirtschaftlicher Schlagwortlisten in das Simple Knowledge Organization System (SKOS)

Christian Rüh, Jens Wiebensohn

Agrar- und Umweltwissenschaftliche Fakultät
Universität Rostock
Justus-von-Liebig-Weg 6
18059 Rostock
christian.rueh@uni-rostock.de
jens.wiebensohn@uni-rostock.de

Abstract: Moderne Farm Management Informationssysteme (FMIS) können durch die Integration von Geo-Webdiensten erheblich an Flexibilität und Aussagekraft gewinnen. Beim Aufbau überregionaler Geodateninfrastrukturen geht es nicht nur um die Einbeziehung der Metadaten, sondern auch um Wortlisten, also kontrollierte Vokabulare und Thesauri, die unter anderem der Verschlagwortung und der Suche nach Metadaten dienen, da somit einheitliche Suchterme genutzt werden, die die Suche verbessern. Technische Grundlage für solche Thesauri kann u.a. SKOS (Simple Knowledge Organization System) sein (z.B. AGROVOC). In diesem Beitrag geht es darum, den Aufbau und die Funktionsweise eines Programmes mit grafischer Benutzeroberfläche zu beschreiben, mit denen Wortlisten aus typischen Anwendungsfällen (Düngung, Pflanzenschutz) von einer CSV Datei in das SKOS Format überführt werden und anschließend in einem RDF Format wie N-Triples abgelegt werden können, das sich dann problemlos z.B. in iQvoc importieren lässt. Darauf aufbauend können z.B. Metadaten von landwirtschaftlicher Geodaten verschlagwortet werden und als Grundlage für eine weitergehende Nutzung von Technologien des Semantic Web dienen.

1 Einleitung

Moderne Farm Management Informationssysteme (FMIS) werden in Zukunft verstärkt mit Daten und Diensten aus (Geo)dateninfrastrukturen konfrontiert werden, wie aktuelle Entwicklungen bei der internetgestützten Bereitstellung von Geodaten mit Relevanz für den Agrarsektor (INSPIRE, ALKIS) zeigen. Dies bedeutet jedoch auch, dass häufig verschiedenartige Begriffswelten aus unterschiedlichen Bereichen aufeinandertreffen. Der mit der Schaffung von Geodateninfrastrukturen einhergehende Bedarf an Datenintegration erfordert ein gemeinsames Verständnis der verwendeten Begriffe. Diese Begriffe sind häufig implizites Wissen der Akteure bzw. werden ohne explizites Vokabular in unterschiedlichen Datenmodellen und Anwendungen genutzt.

Bereits bestehende Vokabulare treten zudem häufig nur einsprachig und in unterschied-

lichen elektronischen Formaten (xls, pdf) auf, was den Austausch der Begriffe auf internationaler Ebene zusätzlich erschwert. Häufig zitiert wird der multilinguale Thesaurus der FAO (AGROVOC), welcher jedoch bei Weitem nicht alle Begriffe des Farm Managements beinhaltet und dessen Pflege bei einer einzelnen Institution angesiedelt ist. Grundsätzlich kann ein solches Vokabular jedoch genutzt werden, um es in formale Ontologien zu überführen [KJ12]. Die Bereitstellung von kontrollierten Vokabularen (z.B. Schlagwortlisten) im World Wide Web stellt einen einfachen Einstieg in das Semantic Web dar [IE09]. Neben AGROVOC¹ sind den Autoren nur wenige weitere Quellen² bekannt, die landwirtschaftliche Schlagwörter auf Basis von RDF veröffentlichen. Die Notwendigkeit des Einsatzes von kontrollierten Vokabularen für automatisierte landwirtschaftliche Anwendungen stellt [NaS09] dar. Ein weiteres Anwendungsfeld ist eine Vereinfachung der Dokumentationspflichten von Landwirten, z.B. bei Agrarumweltmaßnahmen [OA11]. Verfügbare maschinell lesbare Schlagwortlisten mit Relevanz für den landwirtschaftlichen Sektor gibt es beispielsweise mit den Thesauri GEMET³ und UMTES⁴ im Umweltbereich. Ein allgemein anerkanntes (domänenübergreifendes) Vokabular zur Beschreibung von kontrollierten Vokabularen im Web stellt SKOS bereit und seine wichtigsten Elemente werden im folgenden Abschnitt kurz erläutert.

2 SKOS

SKOS (Simple Knowledge Organization System) ist eine semi-formale Repräsentationssprache für Wissen über Vokabulare innerhalb des Semantic Web. Ziel dieser Spezifikation des W3C ist die Bereitstellung eines einfachen Modells zur Beschreibung kontrollierter Vokabulare im Web. Häufig können bestehende Vokabulare wiederverwendet bzw. ergänzt werden, was den Aufwand weiter reduziert. Aus den genannten Gründen eignet es sich in besonderer Weise für einen kostengünstigen Einstieg in das Semantic Web und kann als Brückentechnologie zwischen der informellen Welt (z.B. soziale Netzwerke) und rein formalen Ontologiesprachen (z.B. W3C OWL) dienen [IE09]. Das eigentliche Vokabular der SKOS-Spezifikation⁵ selbst enthält 32 Elemente, welche neben der Klasse *skos:Concept* in die Gruppen *Concept Schemes*, *Lexical Labels*, *Notations*, *Documentation Properties*, *Semantic Relations*, *Concept Collections* und *Mapping Properties* aufgeteilt werden können. Die Klasse *skos:Concept* dient zur Beschreibung eines Begriffs oder einer Vorstellung, welche sich innerhalb eines *skos:ConceptScheme* befindet. Die *Lexical Labels* erlauben die Definition verschiedener Bezeichner aus verschiedenen natürlichen Sprachen für ein *skos:Concept*.

Weiterhin besteht die Möglichkeit, zusammengehörige *skos:Concept(s)* in *skos:Collection* oder *skos:OrderedCollection* zusammenzufassen. Eine Untergruppe der *skos:semanticRelations* sind die *skos:mappingRelations*. Sie dienen zum Ausdruck des Mappings und Verlinkens von *skos:Concepts* in unterschiedlichen

¹ <http://aims.fao.org/standards/agrovoc/about>

² <http://data.igreen-services.com/agrordf>

³ <http://www.eionet.europa.eu/gemet/>

⁴ http://data.uba.de/umt/de/hierarchical_concepts.html

⁵ <http://www.w3.org/TR/skos-reference/#vocab>

skos:ConceptSchemes. Für weitergehende Informationen zum SKOS-Datenmodell wird auf die Referenz verwiesen.

3 Veröffentlichung bestehender Schlagwortlisten mit CVS2SKOS

Wie einleitend bereits ausgeführt, existiert eine große Anzahl von Schlagwortlisten bereits in verschiedenen elektronischen Formaten (xls, pdf). Um diese Ressourcen auf eine einfache Weise für die weitere Nutzung im Internet verfügbar zu machen, kann man sich Konvertierungstools (z.B. CVS2SKOS) zu Nutze machen. Folgende Schritte werden bei der Nutzung von CVS2SKOS durchlaufen [Rü12]:

- 1. Einlesen der Wortliste
- 2. Zuordnung der Spalten zu SKOS Eigenschaften (s. Abb. 1)
- 3. Aufsetzen des Dokument
- 4. Füllen des Dokumentes mit Begriffen/Konzepten
- 5. Matchen der Begriffe mit anderen SKOS Schemas
- 6. Abspeichern der XML RDF Struktur im XML, Turtle bzw. N-Triples Format

Abbildung 1 zeigt das Zuordnen der SKOS Eigenschaften mit Hilfe einer grafischen Benutzeroberfläche (Schritt 2):



Abbildung 1: Zuordnung der Spalten einer CVS-Datei zu SKOS Eigenschaften

Nach erfolgreichem Durchlauf der weiteren Schritte erhält der Nutzer ein standardkonformes SKOS Dokument (Listing 1).

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix cl: <http://croplis.ex/>.
@prefix coll: <http://croplis.ex/collections/>.
@prefix schema: <http://croplis.ex/schema#>.
:#Crop a skos:Concept;
skos:prefLabel "Nutzpflanze"@de;
```

```

skos:altLabel "Crop"@en;
skos:narrower :#Crop_rotation;
skos:narrower :#Crop_Variety;
dct:creator "Vorname Nachname";
dct:modified "2012-10-08T14:03:38+01:00";
skos:broadMatch <http://aims.fao.org/aos/agrovoc/c_5993>;
skos:closeMatch <http://aims.fao.org/aos/agrovoc/c_1972>.

```

Listing 1: Auszug aus einer Schlagwortliste mit SKOS Annotationen (Turtle-Syntax)

4 Zusammenfassung

Mithilfe des vorgestellten Werkzeugs ist es gelungen - unter ausschließlicher Verwendung freier und Open Source Software - landwirtschaftliche Schlagwortlisten im Excel- bzw. CSV-Format in das SKOS-Format zu überführen, optional auch mit Semantik anzureichern (Anbindung an andere Ontologien usw.) und in iQvoc zu importieren. Anschließend können Experten Begriffe um Übersetzungen, Definitionen usw. erweitern und verschiedene landwirtschaftliche Schlagwortlisten zusammenführen.

Literaturverzeichnis

- [IE09] Isaac, A., Ed, S. (2009): SKOS Simple Knowledge Organization System Primer. W3C Working Group Note 18 August 2009. <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/> (25.10.2012).
- [KJ12] Kless, D., Jansen, L., Lindenthal, J., Wiebensohn, J. (2012). A method for re-engineering a thesaurus into an ontology. In: M. Donnelly & G. Guizzardi (Hg.): Formal Ontology in Information Systems. *Frontiers in Artificial Intelligence and Applications*, 133–146.
- [NaS09] Nash, E. (2009). The need for content-lists, dictionaries and ontologies in expressing and evaluating compliance to crop-production regulations, guidelines and standards. In: R. Bill, P. Korduan & L. Theuvsen (Hg.): Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung. Referate der 29. GIL Jahrestagung 09.-10. März 2009, Rostock, Bonn: Köllen Druck+Verlag GmbH, 121–124.
- [OA11] Oenema, O., Amon, B., van Beek, C., Hutchings, N., Perez-Soba, M., Procter, C., Pietrzak, S., Velthof, G.L., Vinther, F., Wilson, L. (2011). Farm data needed for agri-environmental reporting. Technical document summarizing the findings of the DireDate project for the Final Seminar in Luxembourg on 28 March 2011. Publications Office of the European Union,; Luxembourg.
- [Rü12] Rüh, C. (2012). Überführung von Schlagwortlisten in SKOS. In: S. Weisbrich, R. Kaden (Hg.): Entwicklerforum Geodäsie und Geoinformationstechnik 2012. Junge Wissenschaftler forschen, im Druck.