

Universität Leipzig
Fakultät für Mathematik und Informatik
Wintersemester 2021/22
Bachelorarbeit

Transformation von Daten des Amtlichen Liegenschaftskatasterinfor- mationssystem in semantische Datenformate

Betreuer: Prof. Dr. Hans Gert Gräbe, Dr. Christian Zinke-Wehlmann, Norman Radke

Autor: Robin Seidel

Abgabedatum: .03.2022

Inhaltsverzeichnis

1	Einleitung	1
1.1	Themenbeschreibung	1
1.2	Motivation	1
2	Grundlagen	2
2.1	ALKIS	2
2.2	ALKIS DATA	2
2.3	RDF/RML	3
2.4	RDF vs XML	4
2.5	RDF Schema, Ontologies	4
2.6	RML/Yarrrml	4
2.7	Abkürzungen	4
3	State of the Art	5
3.1	Problembeschreibung	5
3.2	Lösungsmöglichkeiten	5
4	Meine Implementierung	5
4.1	genutzte Technologien	5
4.2	Funktionsweise	5
5	Ergebnisse und Evaluation	6
6	Ausblick	6
7	inhalte weiß noch nicht wo hin	6
8	Quellen	6

1 Einleitung

1.1 Themenbeschreibung

1.2 Motivation

Mit der Einführung des World Wide Webs im Jahre 1990 durch Tim Burner Lee ist das Internet heute viel mehr als das, was es früher einmal war. Der Mensch nutzt das Internet immer häufiger als Hauptinformations- und Interaktionsmöglichkeit über diverse Unterhaltungskanäle wie Facebook, Youtube oder andere Websites. (QUELLE) Während am Anfang lediglich ein paar Dutzend Webseiten existierten sind es mittlerweile über zwei Milliarden mit 100 Milliarden Webdokumenten, die sich wiederum alle 6 Monate verdoppeln. (vgl. Meinel 2021) Aber es entstehen nicht immer nur mehr Webseiten auch die Anzahl der Internetnutzer hat sich in den letzten Jahren drastisch gesteigert. 1995 hat der Spiegel 250.000 Nutzer in Deutschland geschätzt, während 2018 durch das statistische Bundesamt ein Anstieg auf 90 wurde. (vgl. Köllinger 2003 zitiert nach: Der Tagesspiegel vom 19. August 1995, S. 23) Demnach nutzen mittlerweile 66,5 Millionen Menschen ab 10 Jahren das Internet als Informations- und Interaktionsquelle. (vgl. Statistisches Bundesamt 2018) Bei einer solchen Masse an Informationen und Daten wird es immer schwerer relevante Informationen herauszufiltern, obwohl das Hauptziel des Internets einst nicht die Unterhaltung, sondern die Informationsbeschaffung war. (QUELLE) Deshalb ist es wichtig die vorhandenen Daten mit Hilfe von RDF Graphen in einer geordneten Art und Weise zur Verfügung zu stellen. Das semantische Web war bereits eine Idee des Entwicklers/Gründers Tim Burner Lee. Er hatte die Vision, dass das Netz von Informationen nicht nur Dokumente und Daten miteinander verbindet, sondern das diesen Daten auch eine Bedeutung gegeben wird. Damit ermöglicht man einer Maschine die Anhäufung von Daten automatisiert nutzen zu können.

- beschreibung semantik web und wichtigste Prinzipien - 4 prinzipien: tim burner lee - nutzen von URI's als namen für objekte - zur eindeutigen identifikation für dinge - nutze HTTP URIs - so können Menschen die Objekte nachschauen - nutze anerkannte standarts(RDF, Sparql) - linke URIS weiter Das semantische Web ist eine Art rohe(also unearbeitete) Daten zuübertragen. (Interoperabilität) Um die Informationen die das Web liefert in einen strukturierten Weg wiederzugeben. Mit Hilfe der daraus entstehenden strukturierten Daten kann die Software diese Daten anschließend besser nutzen und verarbeiten. Das semantische Web ist damit sozusagen eine Alternative zu Webseiten die ausschließlich für die menschliche Nutzung geeignet sind. Die Vorteile die eine Semantifizierung der Daten mit sich bringt sind unter anderem das Abfragen durch Sparql Querys, eine Verbesserung der Bedeutung der Daten, das Verbinden von Daten und die Nutzbarkeit für andere Bereiche.

- überleitung auf Alkis Thematik Alkis Daten(siehe Begriffsklärungen) werden zwar in gewissen Standart und Einheitlichkeit erhoben, allerdings ist die Verarbeitung dieser

Daten hauptsächlich für GeoProgramme(näher erklären) vorgesehen. Die Art der Daten ist nicht für die optimale Verarbeitung von ComputerProgrammen und zur Nutzung für den gewöhnlichen Nutzer vorgesehen. Eine Überführung in Rdf Graphen ermöglicht es das Potenzial der Daten freizuschalten und die Nutzung für zahlreiche andere Bereiche zur Verfügung zu stellen. Quellen: (<https://www.spektrum.de/kolumne/semantic-web-wie-das-internet-inhalte-verstehen-koennte/> 1841389) paper über tim terner lees arbeit https://www.researchgate.net/publication/307845029_Tim_Berners-Lee_s_Semantic_Web paper von burnier über das semantic web 10.3.2022 <https://www.w3.org/2000/Talks/0906-xmlweb-tbl/text.htm> statistik bundesamt https://www.destatis.de/DE/Presse/Pressemitteilungen/2018/09/PD18_330_634.html Köllinger, Philipp <https://www.econstor.eu/handle/10419/151237> (Verlag und Jahr noch angeben)

2 Grundlagen

2.1 ALKIS

ALKIS - das Amtliche Liegenschaftskatasterinformationssystem. - ADV verwaltet ALKIS - ADV hat grundlegende Rahmenbedingungen beschlossen: verwendung internationaler standards, einheitliche objektabbildungskataloge, einheitliche dateninhalte, einheitliche datenaustauschformate, onlinefähigkeit (quelle s 10) - seit 2015 in allen Bundesländern aktiv - vor Alkis: Daten über Automatisches Liegenschaftsbuch(text) und aut. Liegenschaftskarte(graphischer teil) - Unterschiedliche Datenformate, teilweise technisch veraltete Konzepte, getrennte und teilweise redundante Datenhaltung, länderspezifische Implementierung (Quelle s8 13.3) -> vorteile von alkis gegenüber vorgängersystem(seite 11)

- Was sind Kataster - und Liegenschaftsinformationen? - Sind Geobasisdaten über Liegenschaften(Flurstücke und Gebäude) - werden durch Vermessungs und Katasterverwaltungen der Länder erhoben (eventuell nähere Beschreibung notwendig) /url<https://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/ALKIS/broker.jsp?uMen=e5670f15-8e71-3c01-e1f3-351ec0023010>

- durch Alkis werden alle daten des Liegenschaftskatasters redundanzfrei zusammengeführt. Somit wurde eine Einheitlichkeit erschaffen mit internationalen Standards

<https://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/ALKIS/> 11.3.22 ADV Online - ALkis Beschreibung

2.2 ALKIS DATA

https://www.lgln.niedersachsen.de/download/126790/Basiswissen_

[ALKIS_ETRS89_Schulungsmaterial_Stand_12.04.2010.pdf](#)

Alkis Daten, kommen aus Liegenschaftskarte, Liegenschaftsbuch und Punktnahweis

Daten bekommt man von Normbasierter Austauschschnittstelle (NAS) - beruht auf internationalen Standards und Normen - Daten werden im XML Format bereit gestellt - nach dem Grundsatz so einfach wie möglich und so komplex wie nötig
Seite 87 - wichtigstes Feature AX-Flurstueck: enthält folgende Daten(noch ergänzen) - alkis flurstücksdaten werden im Rahmen der Open- Data Lizenz "Datenlizenz Deutschland - Namensnennung-Version 2.0" bereitgestellt. - Abruf der Daten von Schnittstellen der einzelnen Bundesländer - Integration der Schnittstellen unterscheidet sich (Zugriffsrechte etc..) - siehe Quellen Tabelle (Ausschnitt einfügen) Unterscheidung von Geodatendiensten Es wird generell zwischen einem WMS, Web Map Service und zwischen einem WFS, Web Feature Service unterschieden. Der erstere dient zur Visualisierung von Geodaten. Grundlage dafür sind Vektor und Rasterdaten. Ergebnis einer Anfrage auf einen WMS ist ein Kartenausschnitt. Beim WFS bleibt die volle Struktur der Daten erhalten und die Daten werden in vollem Ausmaß bereit gestellt. Und sind somit zur Weiterverarbeitung geeignet. Ist ein Web Coverage Service. Raster und Vektordaten - Daten werden vierteljährig aktualisiert WFS hat verschiedene Abfrageoptionen: GetCapabilities Abruf von Fähigkeiten und Metadaten des aufgerufen Dienstes DescribeFeatureType abfrage von struktur eines featuretypes GetFeature zugriff auf daten <https://onmaps.de/onmapsflurkarte.html>

https://www.bezreg-koeln.nrw.de/brk_internet/geobasis/webdienste/anleitung_wfs.pdf

2.3 RDF/RML

- vom W3C erstellt, unterliegt Regeln <https://www.w3.org/TR/rdf-concepts>
Das Resource Description Framework RDF ist ein universelles, maschinenlesbares Format, das für den Austausch und der Darstellung von Daten gedacht ist. Original wurde es genutzt um Metadaten zu beschreiben (Beleg fehlt noch) und später wurde es in vielen verschiedenen Gebieten und generellen Anwendungen benutzt. Die Daten im RDF Format werden als Graph angeordnet. Ein Graph besitzt Kanten und Knoten. Ein RDF Graph ist eine Ansammlung von Triples, Triples sind Aussagen über Dinge. Ein Triple besteht aus einem Subjekt, was konkret eine URI oder ein leerer Knoten sein kann, einem Prädikat was in dem Triple die Eigenschaft widerspiegelt. Weiterhin besteht das Triple aus einem Objekt. Ein Beispiel wäre:

@prefix foaf: <http://xmlns.com/foaf/0.1/> . @prefix ex: <http://example.com/ ex:RobinSeidel
a foaf:Person

ex:RobinSeidel ist das Subjekt, a das Prädikat und foaf:Person das Objekt.

2.4 RDF vs XML

Nun könnte man sich die Frage warum sollte man RDF nutzen, wenn dies ohnehin schon auf XML beruht und die ALKIS Daten schon in diesem universellem Format vorliegen. Was ist der Vorteil von Rdf gegenüber anderen Datenaustauschformaten? <https://www.w3.org/DesignIssues/RDF-XML.html> <https://dbs.uni-leipzig.de/html/seminararbeiten/semSS99/arbeit5/Rdf.html> Eine Aussage, kann in XML durch verschiedene Arten ausgedrückt werden. Für den Menschen der Sie liest bedeuten Sie das Gleiche. Für eine MASchine/Computer sind es verschiedene XML Bäume. ..

2.5 RDF Schema, Ontologies

- noch nicht sicher ob notwendig rdf schema: - beschreiben vocabularies beschreibung mit rdf triples - data über data -> metadata - vocabulary, sammlung von standart formaten zur wiederverwertbarkeit (nampespace) - gespeichert in owls -> !Unterschied zur Objektorientierung - wenn eine instanz eine bestimmte property einer klasse besitzt, gehört diese instanz zur klasse(umkehrung zu oo) - sehr große macht -> man nutzt standarts des semantic webs, indem man zur Metadata eine Eigenschaft zur Ressource hinzufügt, dadurch wird diese ressource zum Member einer Klasse(was zuvor nicht der fall war) - subclasses existieren owl: - erweiterung zu rdf schema - ontologies beschreiben spezifische Domainen, damit sie unter Fachleuten und INteressenten besser verbreitet und genutzt werden können - .owls sind collection von triples - gibt ßemantic webßemantic also bedeutung - man braucht nicht die ganze bedeutung abzuspeichern da bestimmte key wörter den daten bedeutung geben ontology: - an ontology is a formal, explicit specification of a shared concept - tom grupber 1992 - formale beschreibungssprache um termini einer domäne und deren beziehungen formal zu beschreiben, sodass software diese versteht

2.6 RML/Yarrrml

- RDF Mapping Language - regeln zum überführen von verschiedenen datenquellen in rdf

2.7 Abkürzungen

- ALKIS: Amtliches Liegenschaftskatasterinformationssystem
- RML: RDF Mapping Language
- RDF: Resource Description Framework
- XML: eXtensible markup language

- ADV: Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland
- URL: Uniform Resource Locator
- URI: Uniform Resource Identifier
- NAS: Normbasierter Austauschchnittstelle

3 State of the Art

3.1 Problembeschreibung

- extract, transform, load problem - die einzelnen Schritte beschreiben

3.2 Lösungsmöglichkeiten

- verschiedene Lösungsmöglichkeiten beschreiben - was ist meine Herangehensweise - Probleme die auftreten beschreiben

4 Meine Implementierung

- zur Bearbeitung des Projektes wurde die Programmiersprache Python eingesetzt - mit den bibliotheken ... - beschreibung - suchmaske - profilanfrage muss schon vorhanden sein
- prototypisch entwickelt - konsolenanwendung -> Daten sammeln(Python API) - api calls an WMS - Daten in RML Format - Umwandlung in lesbare Formate -> Daten Speichern()
- Suchmaschinene fenster (like google) - Abfrage von Informationen mit natürlicher sprache(stadt leipzig) - Ergebnis ist ein Wissensgraph

4.1 genutzte Technologien

RDF Triplestore Was ist eine Graph Store? Welchen store sollte ich nutzen? neo4j vs graphDB neo4j: graph native, index free, fast predictable preformance - highly dynamic <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/> <https://www.youtube.com/watch?v=...> <https://db-engines.com/en/system/GraphDB>

4.2 Funktionsweise

- github mit howto zur selbstaussführung - beschreibung der Transformation mit codeauschnitten

5 Ergebnisse und Evaluation

- ergebnis mit beispielhaften SPARQL abfragen - da alkis data auffindbarkeit und nutzbarkeit mangelhaft - arbeit hat sich zum ziel gemacht dies zu verbessern

6 Ausblick

- weitere Alkis Quellen - benutzermanagment - anreicherung der rdf daten

7 inhalte weiß noch nicht wo hin

howto wfs: https://www.bezreg-koeln.nrw.de/brk_internet/geobasis/webdienste/anleitung_wfs.pdf

Hauptschwierigkeiten des Projektes:

Das Mapping der Rdf Daten. generalisierung der datenquellen und dessen automatisierung der Transformation

8 Quellen

- 4.2.22 [zu](#) einleitungserklärungen