



Domain-specific customization of schema.org based on SHACL

Umutcan Şimşek, Kevin Angele, Elias Kärle, Oleksandra Panasiuk, Dieter Fensel
19th International Semantic Web Conference, 2-6 November 2020

Outline

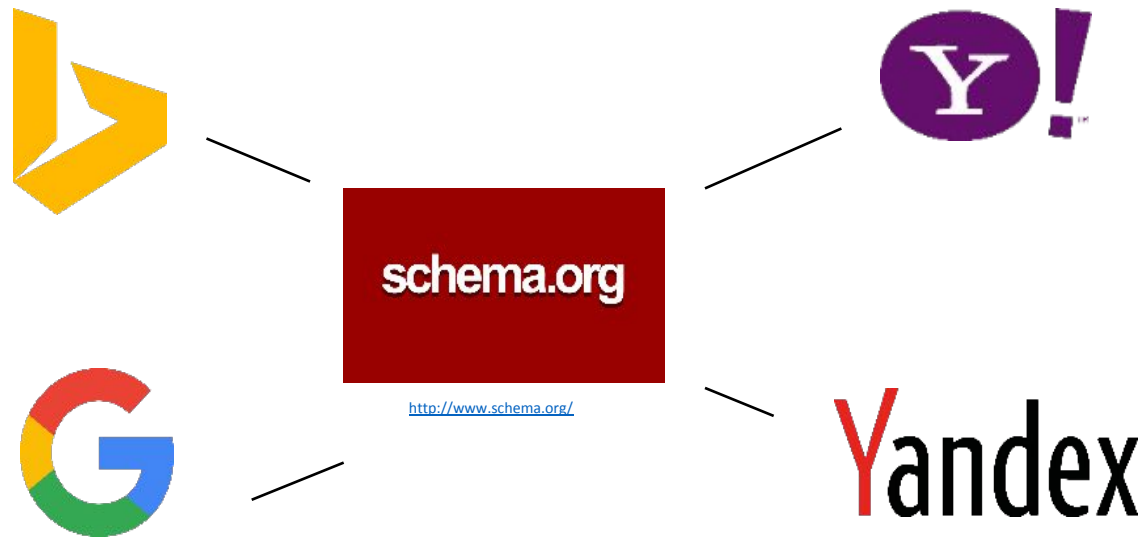
Motivation

Domain Specification Process

Tools

Use Cases

Motivation



De facto industrial standard for annotating web resources

841 types
1369 properties

(numbers in October 2020,
probably more by the time you see
this presentation)

Motivation

The schema.org data model: A gift and a curse...

Covers many domains superficially, not individual domains in detail

“A waterfall can have a phone number”

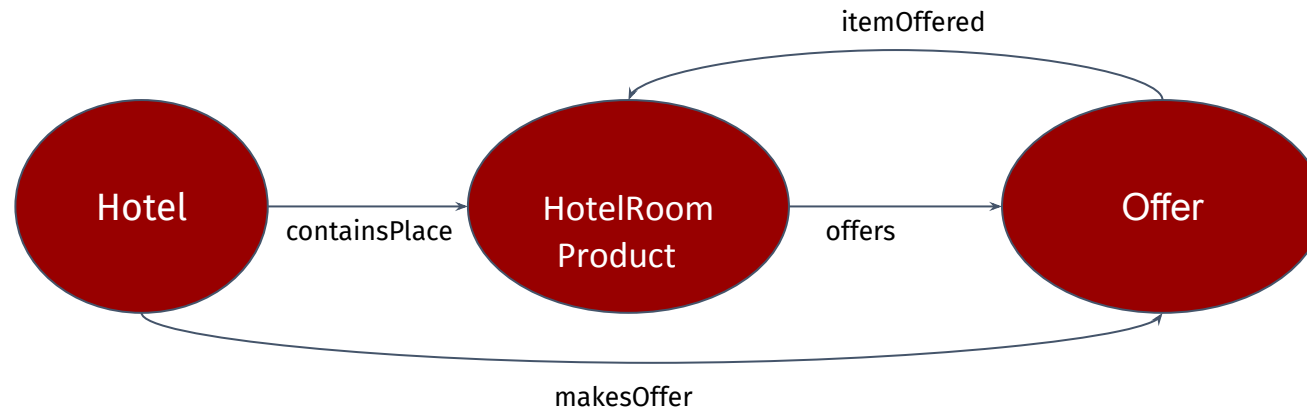
The address information can be represented in at least three different ways

<https://schema.org/docs/datamodel.html>

Motivation

Multi-typed Entities

How to model a hotel, a hotel room and its offers



Motivation

Global vs. local ranges

How do I restrict the range of schema:location property on certain domains?

SportsEvent

Thing > Event > SportsEvent

Event type: Sports event.

location

Place or
PostalAddress or
Text or
VirtualLocation

SportsActivityLocation

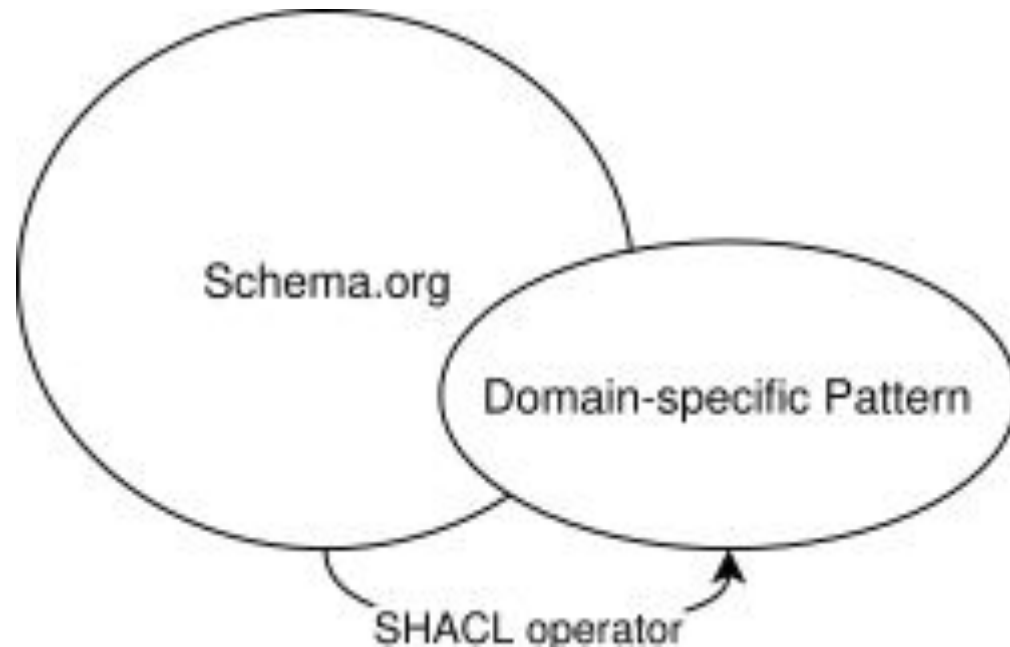
Thing > Organization > LocalBusiness > SportsActivityLocation

Thing > Place > LocalBusiness > SportsActivityLocation

A sports location, such as a playing field.

The location of for example where the event is happening, an organization is located, or where an action takes place.

Domain-specific Patterns

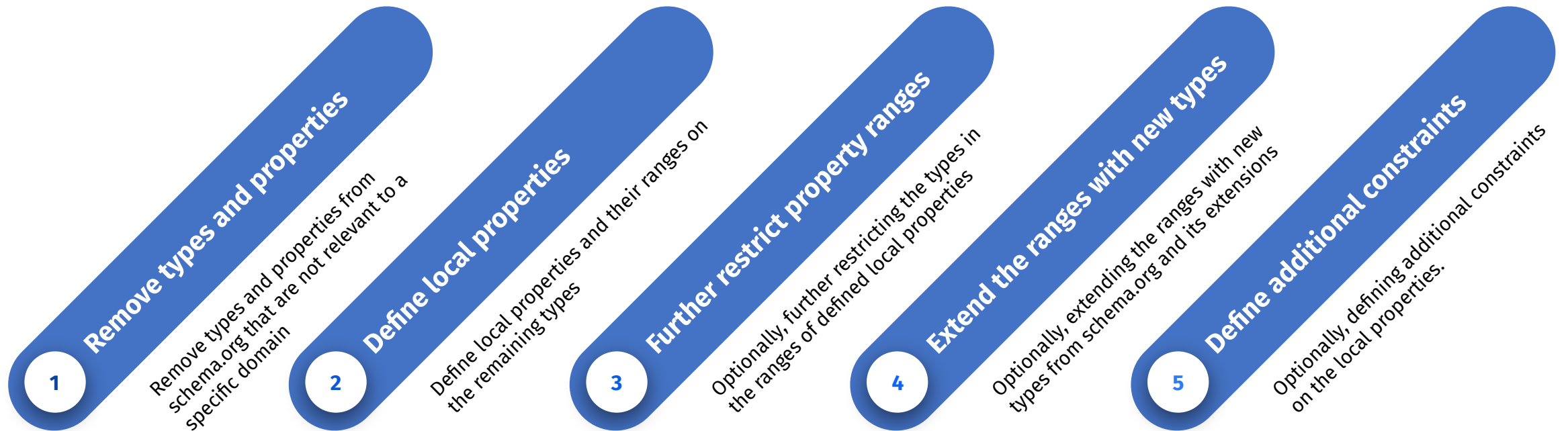


Create extended subsets of schema.org for specific domains

Guide data publishers on their journey of creating semantic annotations

A machine-understandable agreement between domain experts, data publishers and consumers with a subset of SHACL-CORE

Domain Specification Process



Tree View

Show: optional mandatory

Class / Property	Range / Type	Description
└─ LodgingBusiness		A lodging business, such as a motel, hotel, or inn.
├─ name	Text	The name of the item.
├─ location	PostalAddress	The location of for example where the event is happening, an organization is located, or
│ └─ PostalAddress		The mailing address.
├─ checkinTime	DateTime	The earliest someone may check into a lodging establishment.
├─ checkoutTime	DateTime	The latest someone may check out of a lodging establishment.
├─ containsPlace	HotelRoom, Product or n:Sauna	The basic containment relation between a place and another that it contains.
│ └─ HotelRoom, Product		No description found.
│ └─ offers	Offer	An offer to provide this item—for example, an offer to sell a product, rent the DVD of a m
│ │ └─ Offer		An offer to transfer some rights to an item or to provide a service – for example, an offe
│ │ │ └─ name	Text	The name of the item.
│ │ │ └─ description	Text	A description of the item.
│ │ │ └─ priceSpecification	PriceSpecification	One or more detailed price specifications, indicating the unit price and delivery or paym
│ │ │ │ └─ PriceSpecification		A structured value representing a price or price range. Typically, only the subclasses of
│ │ │ │ │ └─ minPrice	Number	The lowest price if the price is a range.
│ │ │ │ │ └─ maxPrice	Number	The highest price if the price is a range.
│ │ │ │ │ └─ priceCurrency	Text	The currency of the price, or a price component when attached to PriceSpecification an
│ │ └─ n:Sauna		a sauna
│ │ │ └─ name	Text	The name of the item.
│ │ │ └─ openingHours	Text	The general opening hours for a business. Opening hours can be specified as a weekly
└─ n:totalNumberOfBeds	Number	total number of beds in an accommodation

SHACL shape: <https://semantify.it/ds/l49vQ318v>

Visualization: <https://semantify.it/domainspecifications/public/l49vQ318v>

Tools

- Part of the **semantify.it** ecosystem
 - Domain Specification Editor
 - Domain Specification Visualizer
 - Annotation Evaluator
 - Annotation Editor

A separate tool demo video is available:
<https://tinyurl.com/yysobz4z>

Edit Domain Specification

Name: Hospitality DS ADVANCED OPTIONS

Description: a DS pattern for lodgingbusinesses

Start Class (1): LodgingBusiness i

+ (Add additional Start Class)

Available Properties

Search for property here

- actionableFeedbackPolicy >
- additionalProperty >
- additionalType >
- address >
- aggregateRating >
- alternateName >
- alumni >
- amenityFeature >
- areaServed >
- audience >

Used Properties

Name	Property Order	Allowed value types	Cardinality
< name	1	<input checked="" type="checkbox"/> Text	<input type="checkbox"/> is optional <input type="checkbox"/> only 1 value
< checkinTime	2	<input checked="" type="checkbox"/> DateTime <input type="checkbox"/> Time	<input type="checkbox"/> is optional <input type="checkbox"/> only 1 value
< checkoutTime	3	<input checked="" type="checkbox"/> DateTime <input type="checkbox"/> Time	<input type="checkbox"/> is optional <input type="checkbox"/> only 1 value

Use Cases

Thüringen Tourism uses schema.org and domain-specific patterns for building their Knowledge Graph and communicating their data model with their IT Solution provider

ThüCAT – Knowledge Graph

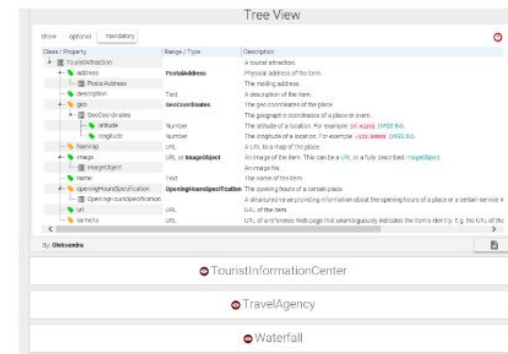
Exkurs semantische Auszeichnung und Struktur

Otto Krause ist am 14.03.1957 in Dresden geboren.

`<firstName>Otto</firstName><familyName>Krause</familyName> ist am <birthDate>14.03.1957</birthDate> in <birthPlace>Dresden</birthPlace> geboren.`



www.semantify.it



Use Cases

DACH-KG / ODTA

A working group that consists of stakeholders in tourism sector from Austria, Germany, South Tyrol and Switzerland

<https://ds.sti2.org>

Organisationen und Forschungseinrichtungen aus Deutschland, Österreich und Südtirol zusammen. Bei den zwei Sitzungen waren Vertreter der Deutschen Zentrale für Tourismus, Österreich Werbung, IDM Südtirol, Tirol Werbung, Vorarlberg Tourismus, Thüringen Tourismus, Tourismus-Marketing Brandenburg, Rheinland-Pfalz Tourismus, AboutCities Niedersachsen, Allgäu, LTS Südtirol und der Hochschule Kempten. Wir tagten am **Semantic Technology Institut an der Uni Innsbruck** und durften nicht nur die Räumlichkeiten nutzen, sondern auch das vorhandene Know How. Elias Kärle mit seinem Team sind inhaltlicher Impulsgeber und stellen Tools zur Verfügung, die die Umsetzungsschritte erleichtern und kontrollieren. Er spricht übrigens auch diese Woche beim Deutschen Tourismustag zu diesem Themenkomplex.



[<- return to DS List](#) [show SHACL serialization](#)

dachkg:Trail

A path, track or unpaved lane or road for sport activities or walking.


[External link](#) [External link to schema.org](#)


Property	Expected Type	Description	Cardinality
identifier	Text URL	The identifier property represents any kind of identifier for any kind of Thing , such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details.	0..N
name	Text	The name of the item.	1
description	Text	A description of the item.	1
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.	0..1
dachkg:startLocation	Place	A sub property of schema.org location . The start location of the trail.	0..1
dachkg:endLocation	Place	A sub property of schema.org location . The final location of the trail.	0..1

Use Cases

German National Tourism Board uses domain-specific patterns define schemas for the data collected from regional tourism organizations to build the German Tourism Knowledge Graph



 **Offenheit** Die lizenzrechtlichen Fragen müssen geklärt werden. Das bedeutet, dass neben Texten auch für Bilder, Videos oder Audiodateien geklärt werden muss, wer welche Rechte woran hat. Sie müssen dann auch entsprechend ausgewiesen werden, damit die weitere Nutzung klar ist.

 **Strukturierung** Daten müssen in einer spezifischen Art vorgehalten werden, damit sie von Maschinen und Menschen interpretiert werden können. Eine im Tourismus etablierte Form der semantischen Auszeichnung ist die nach schema.org und nach seinen erweiterten Domain Specifications, die von der Open Data Tourism Alliance kontinuierlich weiterentwickelt werden.

Die Frage der Auszeichnung von Daten ist gegenwärtig eine der dringlichsten, damit ein einheitlicher Datenstandard etabliert werden kann. Hierzu gibt es eine Dokumentation des erweiterten Vokabulars von schema.org (Domain Specifications) und ein Tool, mit dem die korrekte Auszeichnung von Daten getestet werden kann.

- Übersicht der touristischen Domain Specifications: ds.sti2.org
- Tool zum Testen der strukturierten Daten: www.semantify.it/validator

Preliminary User Studies

- Domain Specification Editor

System Usability Scale (SUS) Survey conducted with 37 participants (28 Tourism students and 9 DACH-KG members)

				Awful		Poor		Good		Excellent	
	\bar{x}	σ	\tilde{x}	\bar{x}	%	\bar{x}	%	\bar{x}	%	\bar{x}	%
DACH-KG	75	25	82.5	-	0	27.5	12.5	77.5	12.5	82.5	75
WSGT-Students	49.82	16.5	50	20	3.4	36.5	17.2	48.23	58.6	70.41	20.7
ALL	55.27	21.24	55	20	2.7	35	16.2	49.86	48.6	76.45	32.4

Table 1. SUS Survey Results

Preliminary User Studies

- Domain Specific Patterns

Survey with 14 computer science students and software developers with some experience with schema.org

Only 21% found domain-specific patterns difficult to understand

All participants reported that domain-specific patterns helped them in some way while creating annotations

Conclusion

- A machine-understandable way to represent domain knowledge for schema.org annotations on the web
- Tool ecosystem provided
- Strong early adoption in tourism

Future Work

- Promote adoption in new domains
- Make larger user studies
- Semi-automatic extraction of domain-specific patterns from Knowledge Graphs



Twitter: @umutsims
<http://umutcan.eu>



www.uibk.ac.at
www.sti2.at