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
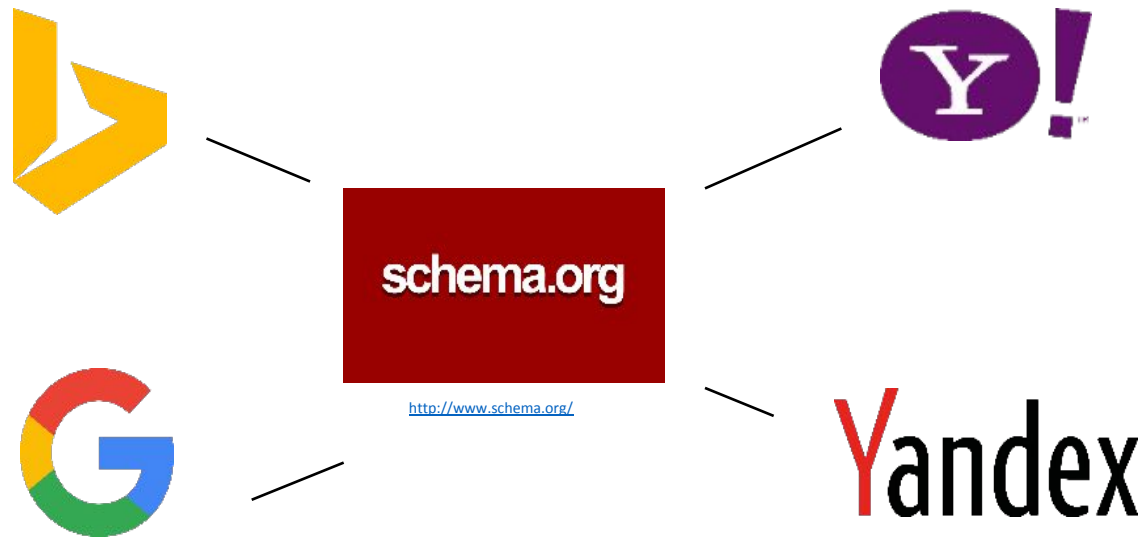
Outline

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)

schema.org

http://www.schema.org/

Motivation

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

Covers many domains superficially, not individual domains in detail
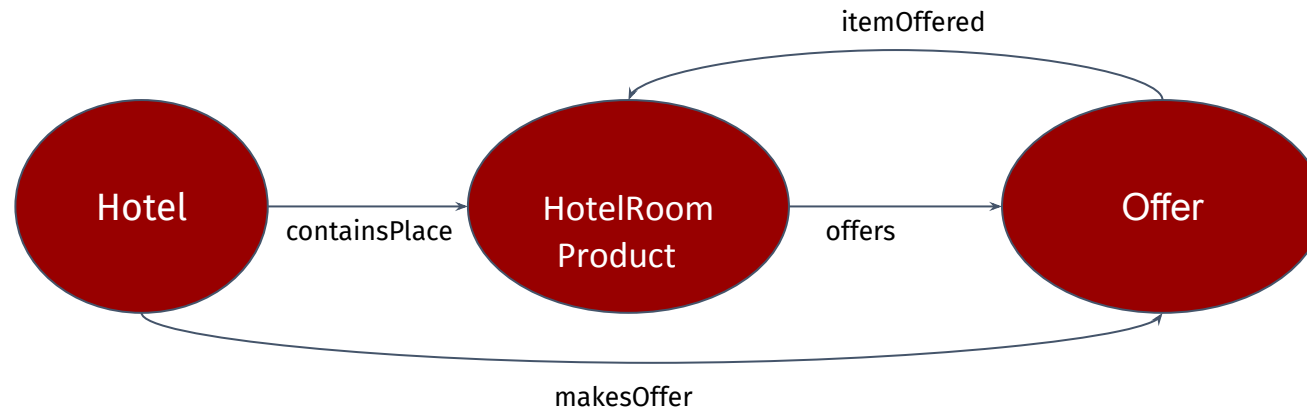
A waterfall can have a phone number

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

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

universität innsbruck

STI INNSBRUCK

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?

Domain-specific Patterns

Schema.org

Domain-specific Pattern
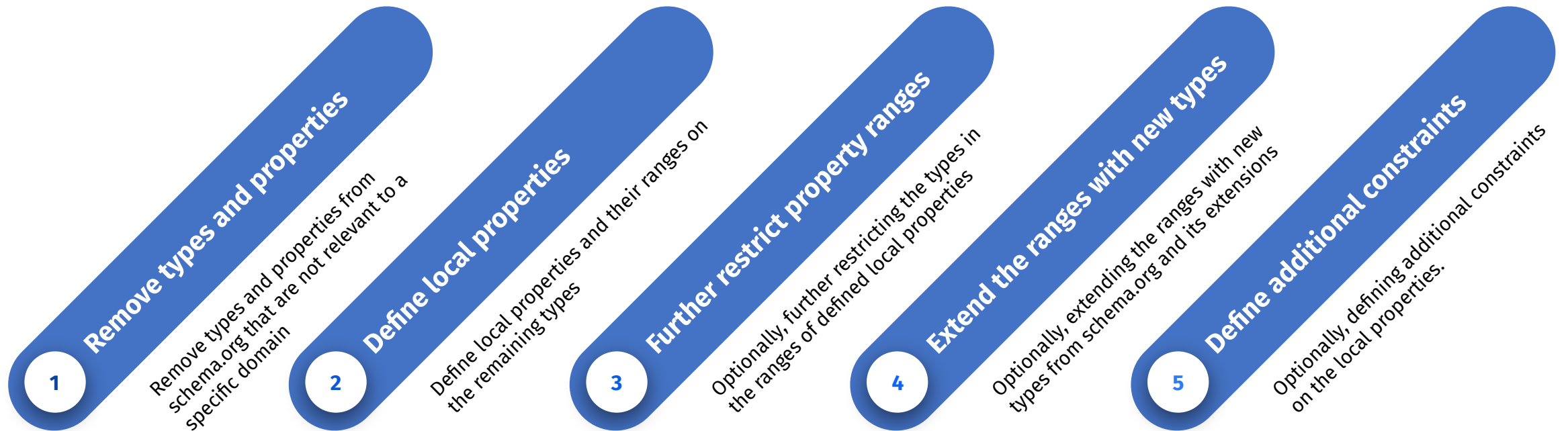
SHACL operator

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

1 **Remove types and properties**
Remove types and properties from schema.org that are not relevant to a specific domain

2 **Define local properties**
Define local properties and their ranges on the remaining types

3 **Further restrict property ranges**
Optionally, further restricting the types in the ranges of defined local properties

4 **Extend the ranges with new types**
Optionally, extending the ranges with new types from schema.org and its extensions

5 **Define additional constraints**
Optionally, defining additional constraints on the local properties.

universität innsbruck

STI INNSBRUCK

Tree View

Show: optional | mandatory |

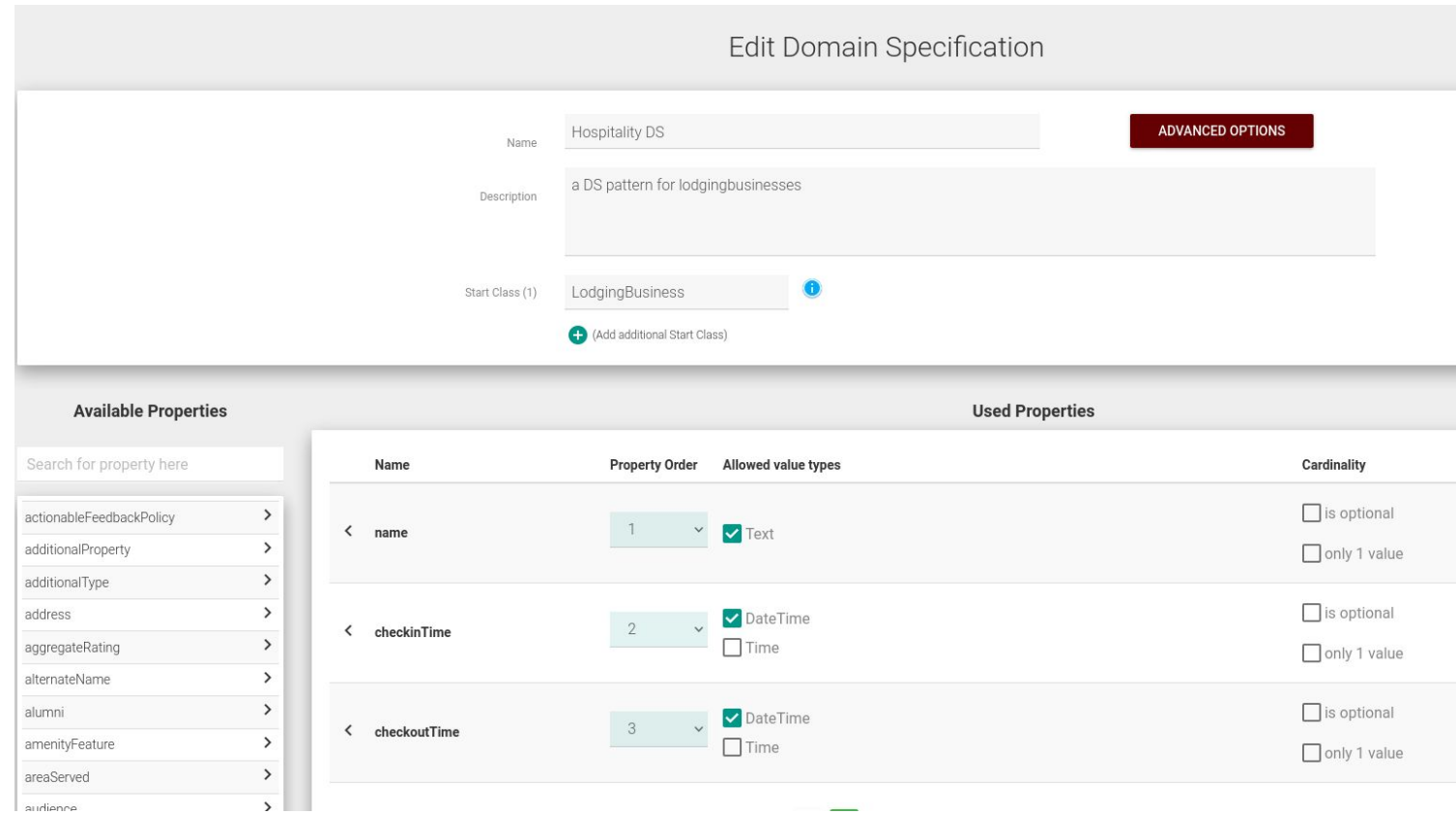| Class / Property	Range / Type	Description
▦ LodgingBusiness		A lodging business, such as a motel, hotel, or inn.
🏷 name	Text	The name of the item.
🏷 location	**PostalAddress**	The location of for example where the event is happening, an organization is located, or
▦ PostalAddress		The mailing address.
🏷 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 off
🏷 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

universität innsbruck

STI INNSBRUCK

Tools

- Part of the **semantify.it** ecosystem

  - Domain Specification Editor

  - Annotation Editor

  - Domain Specification Visualizer

  - Annotation Evaluator

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

Şimşek et al. | ISWC 2020

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

Use Cases

DACH-KG / ODTA
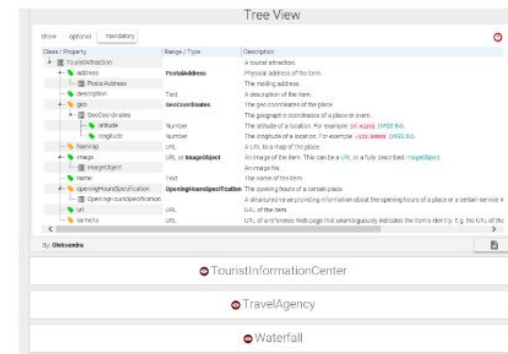
A working group that consists of stakeholders in tourism sector from Austria, Germany, South Tyrol and Switzerland

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 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

universität innsbruck

STI INNSBRUCK

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

Şimşek et al. | ISWC 2020

Preliminary User Studies

- Domain Specification Editor

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

	$\overline{x}$	σ	$\tilde{x}$	Awful		Poor		Good		Excellent	
				$\overline{x}$	%	$\overline{x}$	%	$\overline{x}$	%	$\overline{x}$	%
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

universität innsbruck

STI INNSBRUCK

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

netidee
FÖRDERUNGEN

mindLab

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