



UNIVERSITY
OF APPLIED SCIENCES
UPPER AUSTRIA

Usage of Semantic Web in Austrian Regional Tourism Organizations

SEMANTICS 2019, 11.9.2019

Christina Lohvynenko, Dietmar Nedbal

University of Applied Sciences Upper Austria, Steyr, Austria

HAGENBERG | LINZ | STEYR | WELS

Agenda

Introduction & Motivation

Background and Related Work

Methodology

Discussion of Results

Conclusions

Introduction & Motivation

Tourism as one of the most important economic sectors in Austria:

- Tourism and leisure industry contributed ~16% to the Austrian GDP
- 45 million guests in 2018
- 150 million overnight stays (~65 million in winter season)
 - › domestic overnight stays: 39.4 million, guests from abroad: 110.4 million

International online travel agencies (OTA) dominate the tourism market → Focus on: Websites of Regional Tourism Organizations (RTO)

- Contribute significantly to the promotion of their tourism destinations
- Increase visibility and sales figures on the Internet → semantic annotations

RQ: Has the Semantic Web become a standard in Austria's tourism industry?

Background and Related Work

Studies on quality of content and services offered on official websites of tourism organizations (RTO) with online travel agencies' (OTA) websites

- RTO websites often do not follow state-of-the-art
- OTAs lead in terms of technology usage
 - › (Stavrakantonakis et al. 2014; Kärle et al. 2016; Cao and Yang 2016)

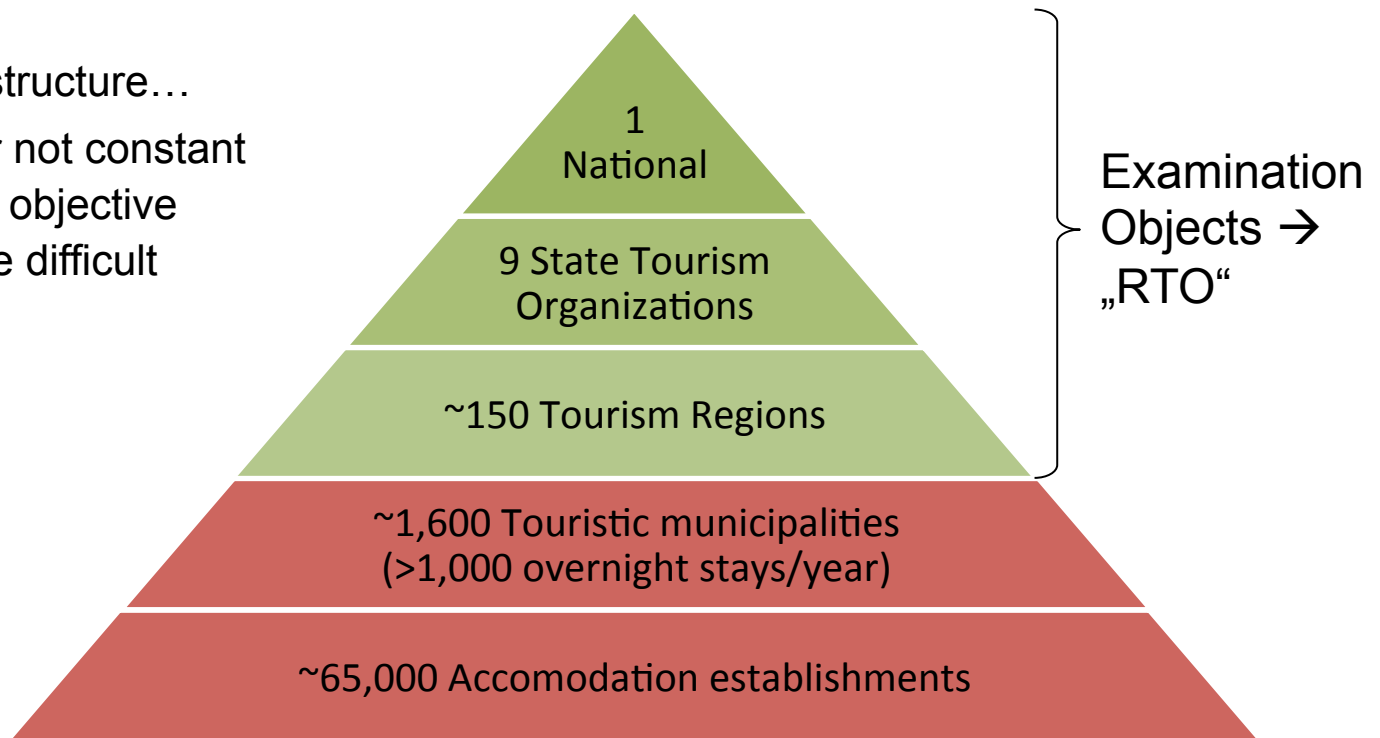
Benefits when using semantic markup

- visibility in the search results of leading search engines (Toma et al. 2014)
- visibility of the promotions being advertised (Fensel et al. 2015)
- enables the use of structured data by emerging intelligent applications (e.g. chatbots/voice search) and improves interoperability among market participants (Hepp et al. 2006; Akbar et al. 2017; Zanker et al. 2009)
- reduce reliance on OTAs

Methodology

Selection of examination objects

- Austrian tourism organizations are well suited as examination objects for this analysis, as they usually have an established website with comparable contents of the region.
- Hierarchical structure...
- BUT: Number not constant which makes objective analysis more difficult



Methodology

Selection of examination objects

- Top-down approach: start with links from national & state tourism org.

National (1) & State Tourism Organizations (9)	Initial list of regional websites	Changes after examination of regional websites	Final examination objects (RTOs)
1 (Austrian National Tourist Office)			1
1 (Burgenland)	3		4
1 (Lower Austria)	6		7
1 (Upper Austria)	26	-2	25
1 (Carinthia)	14	+2	17
1 (Salzburg)	17		18
1 (Styria)	9	+1	11
1 (Tyrol)	35	+5	41
1 (Vorarlberg)	6		7
1 (Vienna)			1
10	117	6	133 websites

Methodology

Data extraction

- Raw web page data: Web Data Commons (Meusel et al. 2014)
 - › data collection “WDC RDFa, Microdata, Embedded JSON-LD, and Microformats Data Sets (November 2017)”
 - › 8,433 files, each around 100mb in size
- Shell script to extract references to the 133 RTOs
 - › One plain text file for each RTO

Preparation of semantic markup

- Create Excel spreadsheets from the text files generated by the shell script using an Excel macro
- Remove duplicated annotations and mentions using filtering rules
- Add columns needed for analysis (RTO, Format, Ontology, Topic,...)

Methodology

Erroneous or incomplete semantic annotations

- No systematic error detection → Semi-automatic (Meusel/Paulheim 2015)
 - Several errors in semantic annotations found
 - › missing slash (e.g. schema.orgPostalAddress),
 - › incorrect cases (e.g. schema.org/webpage/name),
 - › missing data classes (e.g. schema.org/articleBody),
 - › incorrect namespace (e.g. www.schema.org),
 - › incorrect use of properties / values (e.g. schema.org: „price“ in „LodgingBusiness“, vcard: name of tourism organization as „given-name“, wrong date format)
 - Incomplete markup / undefined
 - › <http://www.w3.org/1999/xhtml>
 - › <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
 - › <http://www.w3.org/1999/xhtml/microdata#item>
- } corrected
→ 17.3%
- } not corrected
→ 11.8%

Methodology

Final analysis table

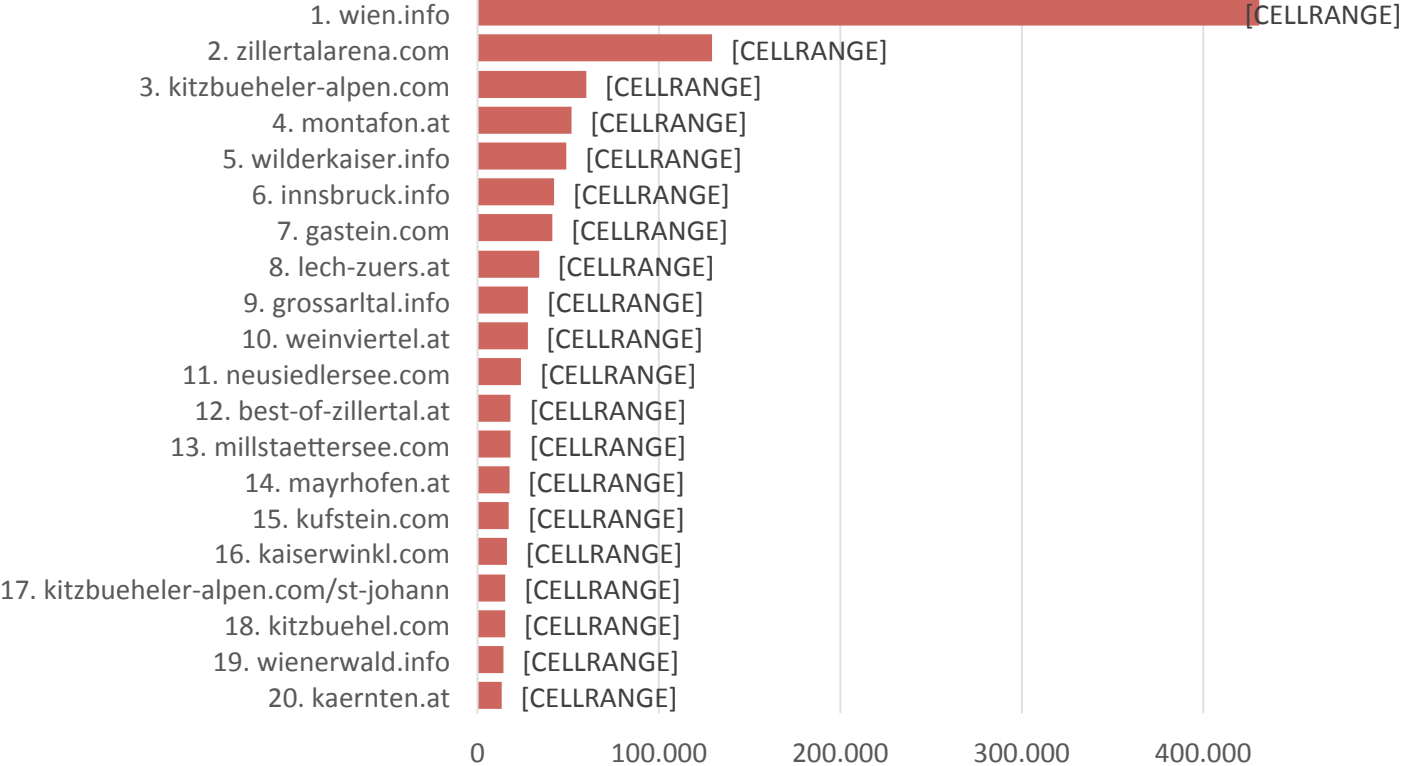
- 430,894 rows (Vienna) and 769,824 rows (77 remaining RTOs w. sem. markup)

	A	B	C	D	E	F	G	H	I	J
	RDF-Quad	Region	Federal	Format	Namespace & data type	Ontology	Class	Property	Topics	Remark
1	_:node111ea2d3486029fbbc554a6f624	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
2	_:node350727337fa5114ba84c51eac24	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
3	_:nodef2ec9c37ddb3bdf93b76bfb172	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
4	_:node6fd21c33d08791d3c495cd7e84	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
5	_:nodeb6b9ea17686dad8334cefb2777	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
6	_:nodeffd294374aebd7f8d43b61a03f1	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
7	_:nodec6f5b613f797285c76e53bc2c96	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
8	_:node4cdee7204a50aa2a2a104d4474	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
9	_:node6f4cdb3e6fe76c119ba41215552	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
10	_:node12d90d613fb0952b1b8b4154b9	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
11	_:node7efff1b027596a794e5c2864483	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
12	_:nodeee835df65c8c88ebfd87a368571f	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
13	_:nodebe3486f11e8689a0d414fedd3b	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
14	_:node2cace2a669e6be952bb53f163a7	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
15	_:nodeb25384f09ee3de45e0551edbe3	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
16	_:node745e32e2a64b87718b67982b8	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
17	_:node4ecaba115f322cd2b2fb8b16458	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
18	_:noded977e4a46a37ee207368c77879	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
19	_:node10aaf5993a43c64a3d7c3d3c59f	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
20	_:nodee0fd38a1fdb33b127d40d26befe	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
21	_:nodec398c42a4bc2d68822deda34d9	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
22	_:nodec1757da5a09efad76eba7248d4	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	
23	_:node16af9d2268fbfd524b151e59ed7	montafon.at	Vorarlberg	Microdata	<http://schema.org/BedAndBreakfast/address>	Schema.org	BedAndBreakfast	address	Product Data	

Discussion of Results

Overview

- 78 Austrian RTOs (59%) use semantic annotations in their websites



Discussion of Results

Formats

- Clear preference of Microdata (93.9%) by number of absolute uses

	RDF QUADS		RTO	
	AMOUNT	PERCENT	AMOUNT	PERCENT
MICRODATA	1,127,678	93.9 %	72	92.3 %
JSON-LD	36,199	3.0 %	19	24.4 %
MICROFORMATS	33,629	2.8 %	20	25.6 %
RDFA	3,212	0.3 %	14	17.9 %

Formats used by Austrian RTOs (n=78)

Discussion of Results

Structured Data Markup: Vocabularies

- Schema.org is preferred, along with its predecessor, Data Vocabulary, in over 80% of all annotations
- Dublin Core terms are used by a large number of RTOs (61 websites) but account to only 3.3% of the overall RDF quads

	RDF quads		RTO	
	AMOUNT	PERCENT	AMOUNT	PERCENT
Schema.org	765,153	63.7 %	63	80.8 %
Data Vocabulary	218,371	18.2 %	18	23.1 %
<undefined>	141,940	11.8 %	68	87.2 %
Dublin Core Terms	39,485	3.3 %	61	78.2 %
vcard	32,062	2.7 %	19	24.4 %
OGP	2,140	0.2 %	11	14.1 %
iCal Schema	1,551	0.1 %	3	3.9 %
XFN	12	0.0 %	1	1.3 %
FOAF	4	0.0 %	1	1.3 %

Austrian RTOs use 8 different ontologies (n=78)

Discussion of Results

Topics

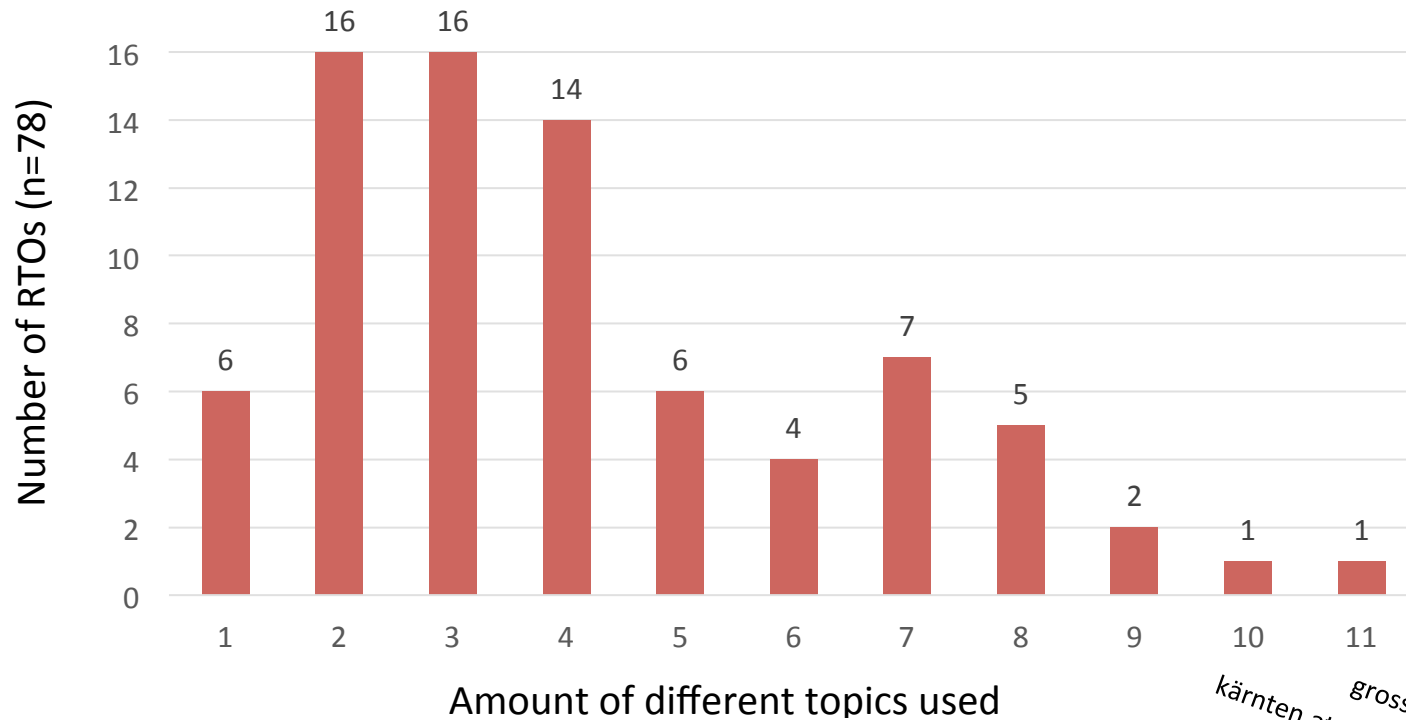
- Groups #1 - #6 from Meusel et al. (2014), remaining groups defined on the basis of the examined data of the RTOs.

#	Topic	Vocabularies and data classes
1	Addresses	s:GeoCoordinates, s:PostalAddress, vcard:Address, vcard:adr, vcard:addressType, vcard:country-name, vcard:email, vcard:locality, vcard:postal-code, vcard:region, vcard:street-address, vcard:tel
2	Blogs	s:Article, s:Blog, s:CreativeWork, s:BlogPosting, vcard:family-name, vcard:fn, vcard:given-name, vcard:n, vcard:Name, vcard:nickname, vcard:note, vcard:title, vcard:url, vcard:vcard
3	Navigational Information	dv:Breadcrumb, s:BreadcrumbList, s:ItemList, s:ListItem, s:url, s:SiteNavigationElement, s:WPFooter, s:WPHeader
4	Organization	dv:Organization, s:Organization, vcard:org, vcard:Organization, vcard:organization-name, vcard:uid
5	People	Foaf:Person, s:JobPosting, s:Person
6	Product Data	s:AggregateOffer, s:AggregateRating, s:Hotel, s:BedAndBreakfast, s:LocationFeatureSpecification, s:LodgingBusiness, s:Offer, s:Product, s:Date, s:PropertyValue, s:Rating, s:Reservation, s:Review, vcard:fn, vcard:n
7	Action	s:SearchAction
8	Event	dv:Event, iCal:component, iCal:description, iCal:dstart, iCal:summary, iCal:vcalender, iCal:Vevent, s:Event, s:Place, vcard:fn, vcard:n, vcard:url, vcard:vcard
9	Images	s:ImageGallery, s:ImageObject, vcard:photo
10	Local Tourism Business	s:Campground, s:GolfCourse, s:LocalBusiness, s:Place, s:TouristAttraction, s:TouristInformationCenter
11	Social Media	dc:source, og:admins, og:app_id, og:description, og:fbmladmins, og:image, og:site_name, og:title, og:type, og:url, s:sameAs, xfn:mePage, xfn:me-hyperlink
12	Website Information	dc:title, s:Language, s:WebPage, s:WebSite

Discussion of Results

Topics

- On average each RTO uses 4.1 different topics.



kärnten.at grossarlal.info

Discussion of Results

Topics

- Use of topics by the 78 RTOs using semantic annotations

	RDF quads		RTOs	
	Amount	Percent	Amount	Percent
Navigational Information	398,947	33.2 %	41	52.6 %
Addresses	176,755	14.7 %	35	44.9 %
Local Tourism Business	134,577	11.2 %	20	25.6 %
Event	94,827	7.9 %	20	25.6 %
Product Data	63,670	5.3 %	24	30.8 %
Website Information	63,130	5.3 %	68	87.2 %
Blogs	52,307	4.4 %	29	37.2 %
Organization	24,182	2.0 %	29	37.2 %
Images	22,301	1.9 %	13	16.7 %
Social Media	21,799	1.8 %	20	25.6 %
Action	4,837	0.4 %	15	19.2 %
People	1,446	0.1 %	10	12.8 %

Discussion of Results

Topics

- Use of topics by the 78 RTOs using semantic annotations

	RDF quads		RTOs	
	Amount	Percent	Amount	Percent
Navigational Information	398,947	33.2 %	41	52.6 %
Addresses	176,755	14.7 %	35	44.9 %
Local Tourism Business				
Event				
Product Data				
Website Information				
Blogs				
Organization				
Images				
Social Media				
Action	4,837	0.4 %	15	19.2 %
People	1,446	0.1 %	10	12.8 %

- Every third semantic markup is made for the purpose of presenting the **breadcrumb** and **list items** that help **navigate** the website.
- **56%** of this topic is annotated using **Schema.org** (dv:Breadcrumb) and **44%** using **Data Vocabulary** (s:SiteNavigationElement).
- **1 RTO** (zillertalarena.com) uses **40%** of all annotations.

Discussion of Results

Topics

- Use of topics by the 7

Navigational Information				
Addresses				
Local Tourism Business				
Event				
Product Data				
Website Information	85,150	5.5 %	68	87.2 %
Blogs	52,307	4.4 %	29	37.2 %
Organization	24,182	2.0 %	29	37.2 %
Images	22,301	1.9 %	13	16.7 %
Social Media	21,799	1.8 %	20	25.6 %
Action	4,837	0.4 %	15	19.2 %
People	1,446	0.1 %	10	12.8 %

- Classes that can be considered as “product” of an RTO (e.g. various types of accommodation). Most used annotations (> 1,000 each) include the LodgingBusiness, AggregateRating, LocationFeatureSpecification, Offer, Hotel, Product, and Review classes.
- RTOs adopted **Schema.org** using **Microdata** format.
- **3 RTOs** (wien.info, montafon.at, kitzbuehel.com) made a total of **91%** of all semantic markup of this topic.

Discussion of Results

Topics

- Use of topics by the 78 RTOs

Navigational Information				
Addresses				
Local Tourism Business				
Event				
Product Data	65,070	5.3 %	24	30.8 %
Website Information	63,130	5.3 %	68	87.2 %
Blogs	52,307	4.4 %	29	37.2 %
Organization	24,182	2.0 %	29	37.2 %
Images	22,301	1.9 %	13	16.7 %
Social Media	21,799	1.8 %	20	25.6 %
Action	4,837	0.4 %	15	19.2 %
People	1,446	0.1 %	10	12.8 %

- Topic describes elements such as the **title**, **alternative names**, **languages** used and individual elements of a website.
- **62%** using **Dublin Core**, the rest by means of **Schema.org**.
- **More than half** of all RDF quads were annotated by **1 RTO** (wien.info)

Discussion of Results

Topics

- With exception of the **three general topics** (“Navigational Information”, “Addresses”, and “Website Information”), the annotation of **RTO’s specific tourism information is strongly influenced by only a few RTOs.**
- No specific touristic annotations found for e.g.
 - › **food establishments** (“FoodEstablishments” class with possible types “Bakery”, “BarOrPub”, “Brewery”, “CafeOrCoffeeShop”, “FastFoodRestaurant”, “IceCreamShop”, “Restaurant”, “Winery”, etc.) or
 - › **ski resorts** (“SportsActivityLocation”, “SkiResort”, etc.), although such content is available on the websites.
- Specific accommodation types are only used by **1 RTO** (montafon.at)
 - › “LodgingBusiness” (w. subtypes “Hostel”, “Hotel”, “Motel”, “Resort”, “Campground”, “BedAndBreakfast”)
- None of the RTOs annotate **specific events** such as “MusicEvent”, “SocialEvent”, “SportsEvent

Conclusions

Summary

- 59% of Austrian RTOs use semantic markup (high ratio in international and industry comparison)
- 20 tourism regions account for 89% of all semantic markup
- Many specific touristic topics that would contribute to unlock the full potential of the Semantic Web are neglected

Limitations

- Findings based on a secondary source (dataset from Nov. 2017)
- No systematic error detection → websites contain a minimum of 29% erroneous or incomplete semantic annotations
 - › Errors bias the analysis results through wrong classification or incorrect detection of semantic markup