




Usage of Semantic Web in Austrian Regional Tourism Organizations

Christina Lohvynenko and Dietmar Nedbal^(✉) 

University of Applied Sciences Upper Austria,
Wehrgrabengasse 1-3, 4400 Steyr, Austria
christina-lohvynenko@gmx.at,
dietmar.nedbal@fh-steyr.at

Abstract. Tourism is one of the most important economic sectors in Austria. Given the high internationality degree of Austrian visitors, the websites of regional tourism organizations (RTOs) are an essential source of information. A state-of-the-art tourism website should include semantic markup for touristic topics so that search engines and other intelligent software applications can access and understand the presented data. This paper empirically studies the usage of Semantic Web formats, ontologies and topics relevant for tourism on the websites of all 137 Austrian RTOs. Results show that 59% of the RTOs use semantic markup. Most regions adhere to the recommendations of leading search engines utilizing ontologies such as Schema.org and the formats Microdata and JSON-LD. While most semantic markup incorporates basic information (e.g. navigation, addresses, corporate data), only few Austrian RTOs annotate touristic relevant topics that would contribute to unlock the full potential of the Semantic Web such as regional events, accommodations, blog posts, images or social media.

Keywords: Semantic Web · Regional tourism organizations · Survey · Austria

1 Introduction

With nearly 45 million resident and non-resident guests in 2018, tourism is one of the most important Austrian economic sectors [1]. In the last years, the tourism and leisure industry contributed around 16% to the Austrian gross domestic product through direct and indirect effects [2]. Even in international comparison, the country occupies an important place among the top 20 tourism destinations [3]. The tourism regions, which are in the midst of the hierarchical organization of this industry in Austria, contribute significantly to the promotion of certain tourism destinations and to addressing a broad target group [4]. These regional tourism organizations (RTO) are also given an important role in the possible weakening of dependence on international online travel agencies (OTA), which dominate the tourism market. Given the growth of the Internet usage and due to the high internationality degree of Austrian visitors, the websites of tourism providers are becoming increasingly important. A state-of-the-art website that implements innovative web technologies is therefore essential [5, 6].

The use of Semantic Web and Linked Data has long been a standard in website optimization and intends to make important content-bearing elements of web pages machine-readable by means of semantic markup so that access to data for search engines and other intelligent software applications is facilitated. The semantic annotation of structured data to a website is one of the most common search engine optimization practices, which is also recommended by leading search engines. Thus, it can increase the online visibility of the web page and the sales figures on the Internet [7–9]. However, the empirical analysis of the use of Semantic Web by the hotel websites in Austria has shown that the use of direct providers in contrast to OTAs is very moderate and often flawed [6, 10]. Such a weak use of structured data in the hotel industry suggests that the Semantic Web has not yet become a standard in Austria’s tourism industry.

With the RTOs playing an important role in the Austrian tourism, the current paper aims to elucidate the usage status of the Semantic Web among these websites. It first discusses the background and related work on the use of structured data in tourism in Sect. 2. Further, the results on an empirical investigation are reported. For this purpose, the selection of the examination objects and preparation of the data for analysis are described in Sect. 3. The results of the evaluation are presented in Sect. 4, followed by a discussion (Sect. 5). Finally, Sect. 6 provides concluding remarks.

2 Background and Related Work

One of the most important communication channels of a tourism organization is the website, which should adhere the current state-of-the-art. In this context it has been recognized that innovative software providing interoperability through ontologies is critical for further innovation in the tourism industry [11]. Although there has been progress in the last ten years, a recent study highlights the still current and growing importance of semantics and ontologies in tourism. The authors further state that academic research in these disciplines is still in its infancy [12].

Website owners and content managers of tourism regions face several challenges when attempting to semantically enrich data on their website. First of all the selection of the appropriate vocabulary, format and content is not a trivial task. In addition to common vocabularies independent of the domain, several domain-specific ontologies for tourism have also been developed which makes it difficult to select the most suitable and, at the same time, a future-proof vocabulary. The Linked Open Vocabularies project, for example, provides a central information point about well-documented vocabularies [13]. The constantly growing website lists 660 high quality vocabularies as of Feb. 2019. Measured by the number of vocabularies that reuse the vocabulary, the most popular ontologies are Dublin Core Metadata Terms (dcterms), Dublin Core Metadata Element Set (dce), Friend of a Friend vocabulary (foaf), A vocabulary for annotating vocabulary descriptions (vann), Simple Knowledge Organization System (skos), Creative Commons Rights Expression Language (cc), SemWeb Vocab Status ontology (vs) and Schema.org vocabulary (schema) [14]. The problem of common vocabularies often lies in the level of precision over domain-specific ontologies. For example, until version 3.0, Schema.org lacked the ability to describe the number of

beds in a room, or whether pets are allowed or not [15]. One of the main goals of tourism-specific vocabularies is to achieve a better interoperability and integration of travel information systems [16]. Several researches have focused on the design of semantic vocabularies for the tourism and travel industry [17] (e.g. Harmonise [18], QALL-ME [19], cDott [20], Accommodation Ontology [21], Tourpedia [22]).

Given the amount and diversity of available ontologies, an industry wide adoption is crucial for a future-proof vocabulary. The Web Data Commons project features the largest publicly available collection of structured data from a non-profit organization [23], allowing researchers to analyze the adoption of structured data across the Web. An analysis for the period 2010 to 2013 showed that the use of the Semantic Web, its formats and data classes has been steadily increasing. The comparison of the 2012 and 2013 datasets revealed that the number of websites using Microdata has even grown by more than factor four in just one year. The topics that received the most attention through semantic markup were people and organizations, blog articles, navigation information, product and event data [23]. In another study focusing only on the adoption of Schema.org, it was shown that about half of the elements of this vocabulary have not been used in any of the websites from the Web Data Commons dataset [24].

Since a website is one of the most important means of communication for tourism organizations, several studies have addressed the quality of touristic websites. International online travel agencies have heavily dominated the tourism sector in recent years. Tourism organizations in Austria are also suffering from this online competition and are trying to counteract this competition by means of innovative technologies and intelligent advertising of products and services on several channels. When comparing the quality of content and services offered on official websites of tourism organizations with online travel agencies' websites, OTA websites have often received better results. Tourism websites often do not follow state-of-the-art online developments, therefore OTAs have the lead in terms of technology usage, according to the studies [6, 10, 25]. As far as Austria is concerned, studies in recent years have distinguished a good performance and numerous innovative integrated services on the websites of official Austrian tourism organizations in international comparison [26, 27].

The use of well-documented structured markup should enable error-free annotation and improve the quality of the website. Unfortunately, a large variety of erroneous and restricted usage in the semantic markup are made in practice when using vocabularies like Schema.org, which hinders real-life applications to use the data [10, 28]. To counteract this problem, Şimşek et al. described an approach that validates Schema.org markup in terms of completeness of the annotations for a specified domain and semantic consistency [29] that was implemented in an online-tool *semantify.it* [30].

Benefits when using Semantic Web technology include better visibility in the search results of leading search engines [7], as well as better online visibility of the promotions being advertised [5]. This further helps reducing reliance on OTAs, enables the use of structured data by emerging intelligent applications (e.g. chatbots and voice search) and improves interoperability among market participants [31–33].

The literature review has shown that the topic of using the Semantic Web has a long history and great potential for the industry. Studies indicate that the tourism sector often lacks expertise and knowledge of the correct use of Semantic Web technology. Furthermore, research on the use of semantic technologies in Austrian tourism organizations

focuses mostly on either the hotel sector or individual tourism organizations. A recent study of the usage of Semantic Web comprising all Austrian regional tourism organizations could not be identified during the literature search.

3 Methodology

The methodology for the empirical investigation started with a definition and selection of the examination objects. This is followed by a description of the data extraction process and the preparation of semantic markup for the actual analysis. It is also detailed, how incomplete and erroneous annotations were identified and how they were assigned to groups that emerged during this analysis.

3.1 Selection of the Examination Objects

Austrian regional tourism organizations are well suited as examination objects for this analysis, as they usually have an established website with comparable contents of the region. However, the number of these organizations is not constant in Austria, which makes objective analysis more difficult.

The organization of Austrian tourism has a hierarchical structure. The basis of the tourism market is provided by the 65,000 tourism businesses, most of which operate in municipalities that are classified as tourism-intensive municipality with at least 1,000 overnight stays per year. Of the 1,568 Austrian tourism-intensive municipalities, 151 were categorized as tourism regions in 2008 [34]. At the state level, tourism in Austria is divided into the respective offices of the nine state governments with one national tourism organization (“Austrian National Tourist Office”) on the top, working closely together with the tourism regions. Therefore, in this work, the tourism regions together with the nine state tourism organizations and the national tourism organization are referenced to as regional tourism organizations (RTO) in the following.

As mentioned, the number of RTOs varies over time. For example, in Upper Austria, a new tourism law came into force, according to which the number of tourism associations (and thus also the RTOs) must be reduced from 100 to 20 by the year 2020. There are tourism associations that have already merged, but still have separate websites (e.g. “Wels” and “Sattledt”) and others that have no joint website (e.g. “Oberes Mühlviertel”) as of June 15, 2018. For this research, the list of RTOs to be examined has been determined in a top-down approach. Starting from the actual references on the nine state tourism organizations websites, an initial list of 117 regional websites was gathered (3 organizations in Burgenland, 6 in Lower Austria, 26 in Upper Austria, 14 in Carinthia, 17 in Salzburg, 9 in Styria, 35 in Tyrol, 6 in Vorarlberg). After examining the individual websites of these 117 organizations, the following changes were made: Two Upper Austrian RTOs without own website (“Nationalpark Region Ennstal” and “Steyrtal”) were removed and RTOs with separate individual websites were added in Carinthia (1 RTO split into 3 websites), Styria (1 RTO split into 2 websites), and Tyrol (2 RTOs split into 7 websites). In total, 133 websites (one national, nine state and 123 regional tourist organization websites) were included, all of which are subsequently referred to as RTO.

3.2 Data Extraction Process

For this research we used data from Web Data Commons [23], making raw web page data, extracted metadata, and snippets of individual web pages available to the public. The data collection entitled “WDC RDFa, Microdata, Embedded JSON-LD, and Microformats Data Sets (November 2017)” was used as basis for data extraction. The original record contains 8,433 files, each around 100 MB in size. The data in the collection is represented in the form of RDF quads with subject, predicate, and object as well as the URL of the web page from which the data was extracted as fourth element.

With the help of a shell script, the downloaded files were unpacked and examined for the presence of semantic annotations of one of the 133 defined RTOs. The script generates plain text files and can be downloaded from the URL <https://t1p.de/shellscript>. The duration of the script was approximately 48 h, with ten tasks run simultaneously on several machines.

3.3 Preparation of Semantic Markup

The preparation of the data for the actual evaluation was done using Microsoft Excel 2016. The first step was to create 133 Excel spreadsheets, one for each tourism region from the text files generated by the shell script using an Excel macro. With the help of conditional formatting, regular expressions, and filtering rules in Excel, duplicated annotations and mentions were removed (repeated use of the same annotation on the same web page) and the markup of all subdomains of the respective RTO were checked and adjusted if necessary. Thus, only those data remained, where the fourth part of the RDF quad contained the domain of one of the 133 RTOs defined.

After all tables were cleaned up with irrelevant data, all individual tables containing structured data were combined in two files (one own Excel file containing “wien.info” markup and one for the remaining 77 websites). This subdivision was necessary due to the limited number of rows in this version of Excel.

In order to be able to identify different types of structured data in websites of Austrian tourism regions, the table has been extended with additional information. The final analysis table can be downloaded from the URL <http://t1p.de/analysistable> as Microsoft Excel file. It contains the following columns:

- The first column contains the relevant RDF quads (430,894 in Vienna and 769,824 in the file for all remaining regions).
- The second column (“Region”) contains the domain of the respective tourism region, gathered from the URL.
- The third column (“Federal State”) allows the assignment to one of the nine federal states and to the national tourism organization of Austria (austriatourism.com).
- The fourth column “Format” contains the format used for a specific semantic markup. This information was taken from the Web Data Commons file name from which the respective RDF quad was extracted (e.g. file “dpef.html-embedded-jsonld.nq” contains the semantic annotations carried out by JSON-LD).

- The “Namespace & data type” column represents the predicate of the respective triples and contains, in addition to the namespace of the ontology, the names of the data classes and data properties used. The namespaces were determined by means of the Excel filter function from the first column containing the RDF quads.
- The “Ontology” column captures the name of the ontology, which was determined by the namespace in the “Namespace & data type” column.
- The “Class” column contains the data classes used and the “Property” column lists the data properties used by the RDF quad. The data on classes and properties was determined using the Excel filter function from the RDF quads themselves or from the “Namespace & data type” column.
- The “Topics” column contains aggregated information of the data classes from various ontologies used into subject areas, containing similar or related objects (cf. Sect. 4.4).
- The last column “Remark” was used to take notes about found errors or incomplete semantic annotations, most of which were previously described in the study of Meusel and Paulheim [28]. Mistakes found include missing slash, incorrect upper or lower case, missing or incorrect use of a data types, incorrect use of namespace, property mapped to an incorrect class or data type, incorrect use of property values, and incomplete/wrong specification of namespace.

4 Analysis Results

This section contains the main findings of the survey on the use of Semantic Web technology by Austrian RTOs. First, an overview of the top 20 RTOs using semantic markup is given. This is followed by a brief analysis of the formats and ontologies used. Finally, insight into the topics that were annotated by the RTOs is provided.

4.1 Amount of RTOs Using Semantic Annotations

A total of 78 Austrian RTOs (59%) use Semantic Web annotations in their websites, while the remaining 55 RTO websites did not show any semantic markup in the course of this analysis.

Figure 1 shows the top 20 RTOs, measured by the absolute number of RDF quads identified. The leading RTO is Vienna (domain: wien.info), which has 430,894 RDF quads integrated into its website. Second place in this ranking is occupied by zillertalarena.com with 129,320 RDF quads. The other 18 RTOs shown in the figure each use more than 10,000 RDF quads. The structured data from wien.info alone make up 36% of the entire data set; zillertalarena.com added another 11% and the remaining 18 RTOs from the top 20 list sum up to 42% of all annotations. The top 20 regions thus make 89% of the total amount of semantic markup.

4.2 Formats

The use of the Semantic Web formats shows a clear preference of the Microdata format (93.9%) by the number of absolute uses in the RDF quads. JSON-LD was used in 3% and microformats in 2.8% of the RDF quads. The use of RDFa is only at 0.3% and includes almost only the Open Graph protocol (OGP).

53.8% of the 78 RTOs with structured data use Microdata as the only format for semantic annotation of website content. The use of multiple formats by RTO is heterogeneous: 10.3% use Microdata and Microformats at the same time, another 9% Microdata and JSON-LD. The three formats Microdata, Microformats and JSON-LD are simultaneously used by 7.7% of the RTOs. RDFa alone is used by four RTOs (5.1%). All four formats are used by three RTOs (3.8%). The remaining 10.3% of the RTOs use a combination of five different formats.

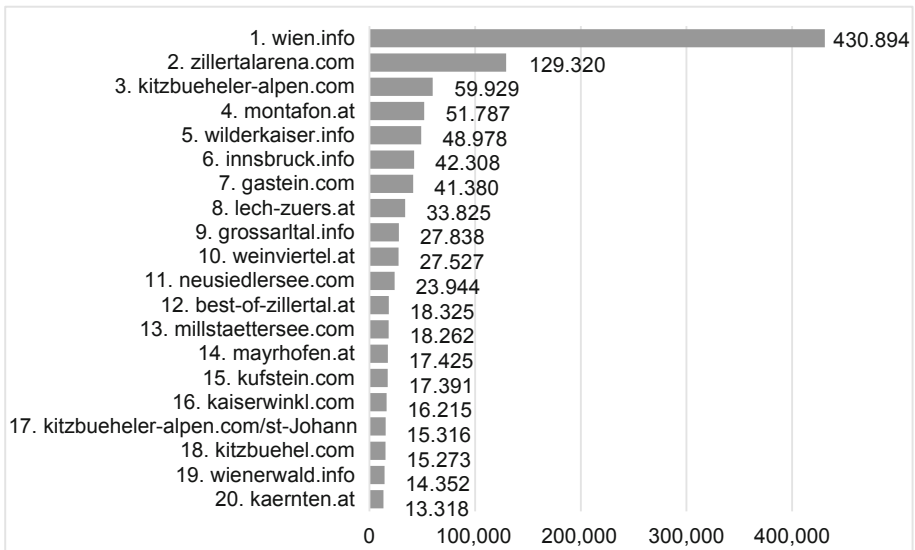


Fig. 1. Top 20 Austrian RTOs by absolute number of RDF quads.

4.3 Structured Data Markup: Ontologies

The examined websites use a total of eight different ontologies. The most used ontology is Schema.org with 63.7% by the number of absolute uses in the RDF quads. In second place (18.2% of the RDF quads) is Data Vocabulary. Dublin Core terms are used by a large number of RTOs (61 websites) but account to only 3.3% of the overall RDF quads. The remaining four ontologies (hCard, OGP, iCal Schema, XFN, FOAF) are all referenced by less than 3% semantic markup. Interestingly, none of the vocabularies developed specifically for tourism were found in the examined objects.

4.4 Topics

Since same or similar content can be annotated using various ontologies and data classes, an overview of the topics that have been covered by the RTOs needs additional consolidation. For this reason, the thematically related objects of a tourism site website were subsequently grouped into similar topics, representing subject areas or categories.

Table 1. Topics and their associated ontologies and data classes.

#	Topic	Ontologies 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

Table 1 presents the twelve topics identified during the analysis, including the list of data classes that make up each group. The first six topics were taken from the study of Meusel et al. [23]. The remaining groups were defined on the basis of the examined data of the RTOs. The ontologies are abbreviated as follows: “s:” stands for Schema.org, “dv:” for Data Vocabulary, “dc:” for Dublin Core, and “og:” for OGP followed by the respective data class.

The subdivision into these twelve topics unfortunately does not guarantee that there is no overlapping in the content. For example, many blog articles contained information on tourist attractions (topic “Local Tourism Business”), pictures in the category “Images” were occasionally identical to the image properties of individual topics such as “Organization”, “Event”, “Local Tourism Business”, or “Blogs” and several classes are also described by properties that contain address information. The Schema.org class “s:Place” has been divided manually into two topics: on the one hand in “Event”, if the information was about an event location, and on the other hand in “Local Tourism Business”. The analysis of the use of topics is presented in Table 2; details on the topics are presented in the following.

Table 2. Use of topics by the 78 RTOs using semantic annotations.

Topic	RDF quads	RTOs
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%)

Navigational Information. Every third semantic markup is made for the purpose of presenting the breadcrumb and list items that help navigate the website. Nearly 56% of this topic is annotated using Schema.org and 44% using Data Vocabulary. Only about 0.1% of the markup is made using JSON-LD and Microdata. A total of eight RTOs account for 81% of the data in the category, of which RTO “zillertalarena.com” alone uses 40% of the annotations. Most commonly used are the classes “dv:Breadcrumb” and “s:SiteNavigationElement”.

Addresses. Almost 15% of the markup contains various address details. The annotations use Schema.org and Microdata in 96% of the cases, the remainder is annotated using the Microformat hCard. 41% of the RTOs annotate address data of the region where the company or local providers are located; the exact address (either street and house number or latitude and longitude) is awarded by 45% of the RTOs. 15% of the RTOs use this topic for specific contact information such as telephone, fax, e-mail or URL.

Local Tourism Business. 11.2% of the RDF quads represent information on this topic. Four RTOs (wien.info, weinviertel.at, innsbruck.info, gastein.com) contribute 84.1% of the data in this topic. The only ontologies used here are Schema.org and Microdata.

Events. Almost 8% of the data represent events in the region. Annotations are made at 98% by means of Schema.org and Microdata, the remainder by the Microformats hCalendar and hCard. The most used property is the start date of an event, followed by the name, image, location, URL, description, address and the special offers. Overall, only two RTOs (wien.info and lech-zuers.at) have made 87.3% of all annotations in this topic.

Product Data. This topic describes both the “Product” and “Offer” data classes as well as various types of accommodation that can be considered as the product of an RTO. 5.3% of all RDF quads found are subsumed under this topic. RTOs adopted Schema.org and Microdata ontologies. Most used annotations (over 1,000 each) include the LodgingBusiness, AggregateRating, LocationFeatureSpecification, Offer, Hotel, Product, and Review classes. Three RTOs (wien.info, montafon.at, kitzbuehel.com) made a total of 91% of all semantic markup of this topic.

Website Information. This topic describes various elements such as the title, alternative names, languages used and individual elements of a website. 62% of the RDF quads were annotated using Dublin Core, the rest by means of Schema.org. The use of Microdata dominated the format use (93%), with JSON-LD making up the remaining 7%. Although 68 RTOs are using this topic, more than half of the RDF quads in this category were annotated by wien.info.

Blogs. In this section, blog, press and web pages published on the website, including author data, titles, descriptions and evaluations, are subsumed. Four regions (best-of-zillertal.at, wien.info, mayrhofen.at and grossarlal.info) out of 29 make 81% of all RDF quads of this topic. Almost half of all annotations are made using Schema.org and Microdata, the rest using hCard. Typical semantic information include headline, description, author name and URL.

Organization. This topic is used to present information about the website operator such as name, logo and VAT number. 96% of the annotations are done using Schema.org, the rest using Data Vocabulary and Microformats. Microdata is used in 69% of annotations, followed by JSON (28%) and the Microformat hCard (4%). The use of this topic is dominated by four RTOs (nationalpark.at, oetztal.com, stantonamarlberg.com and neusiedlersee.com).

Images. This topic contains various pictures and collections of pictures. 99% of the annotations use Schema.org (mainly Microdata), the rest the Microformat hCard. Four RTOs (kaernten.at, kitzbuehel.com, montafon.at and tennengau.com) account to 85% of all annotated images.

Social Media. Social media annotations are made using four different ontologies (primarily OGP and Schema.org, but also Dublin Core and XFN) in all four formats. The most common purpose is to link to the social media presence: 10 RTOs link to

their page on Facebook, five on Instagram, four on YouTube, three on Google+, two on Twitter, and one each on Pinterest and Flickr. Almost 70% of all annotations were made by the RTO neusiedlersee.com.

Action. This topic is used to mark the entries in the search fields or forms that are used by the search engines primarily to provide users with an opportunity to search the content of a website directly on the search results page in their own search window. Four RTOs (grossarlta.info, austriatourism.com, reutte.com and bregenzwald.at) out of 15 account for 91.5% of the markup in this topic, which are made exclusively using JSON-LD and Schema.org.

People. This topic subsumes individuals (article authors, team members, etc.) and company job offers. Most annotations are based on Schema.org and Microdata. Three out of ten RTOs (lech-zuers.at, hoch-koenig.at and mayrhofen.at) make up 94% of all RDF quads in this topic.

5 Discussion

The analysis revealed that the use of Semantic Web in Austrian RTOs complies with the recommendations of leading search engines such as Google, Yahoo, Bing and Yandex. The majority of semantic annotations by tourism regions are made using Microdata and JSON-LD. In addition, considering a total of eight ontologies that are used, the recommended Schema.org is preferred, along with its predecessor, Data Vocabulary, in over 80% of all annotations.

The grouping of semantic markup in twelve thematically related topics allowed an overview of all structured data specifically for Austrian tourism regions - regardless of the formats and ontologies used. The analysis showed that, with the 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. While general information is important to search engines as well as various software agents, specific tourism content should also be semantically annotated to exploit the full potential of the Semantic Web.

For tourism, relevant Schema.org classes and properties are distributed in different parts of this ontology [16]. However, Austrian RTOs use only a few data types and properties of Schema.org intended for the tourism industry. For example, no annotations for food establishments (“FoodEstablishments” class with possible types “Bakery”, “BarOrPub”, “Brewery”, “CafeOrCoffeeShop”, “FastFoodRestaurant”, “IceCreamShop”, “Restaurant”, “Winery”, etc.) or ski resorts (“SportsActivityLocation”, “SkiResort” classes) were found, although such content is available on the websites.

The analysis of the topic “Product Data” revealed that the possibility of specifying specific types of accommodation are hardly used by the RTOs. The Schema.org type “LodgingBusiness” can be used, for example or the more specific subtypes “Hostel”, “Hotel”, “Motel”, “Resort”, “Campground”, or “BedAndBreakfast”. The three types “Hotel”, “Campground” and “BedAndBreakfast” together with the type “LocationFeatureSpecification” are only used by one RTO (montafon.at). Furthermore, none

of the RTOs annotate specific events such as “MusicEvent”, “SocialEvent”, “SportsEvent”, etc. Nevertheless, a precise classification is particularly important for tourism organizations for all available content and such generic classes should be avoided [32].

Detailed information on accommodations that are relevant for a user’s booking decision and also contribute to specific search results (e.g. Schema.org properties like “amenityFeature”, “availability”, “price”, “offer”, “paymentAccepted”, “petsAllowed”, “priceCurrency”, “priceRange”, “availability”) were used by 13 RTOs. Taking a closer look, 92% of RDF quads with such detailed information came from only one region (montafon.at). The remaining twelve RTOs used the properties mentioned only sporadically. As a result, applications need additional data extraction and fusion techniques to understand the content of these sites (e.g. to find out which RTO offers a specific type of accommodation with specific equipment). Thus, the integration of multiple data items representing the same real-world object into a single, consistent, and precise representation remains challenging [9].

6 Conclusion

The present work empirically studies the use of structured data on the websites of Austrian tourism regions. According to the results of this analysis, 59% of the tourism organizations surveyed use the Semantic Web, which is a high ratio in international and industry comparison. However, the use is designed according to the Pareto principle: 20% of the tourism regions account for 82% of all semantic markup. Most tourism regions adhere to the recommendations of the search engines and use the ontology Schema.org and the formats Microdata and JSON-LD. While semantic markup of basic information such as addresses, corporate and website data is necessary, many areas that would contribute to unlock the full potential of the Semantic Web are neglected by Austrian RTOs. The use of touristic relevant topics, such as regional events, accommodations, blog posts, images or social media is dominated by a few RTOs. None of the special tourism ontologies were applied and also only a few classes and properties that are typical for this type of industry are used by a large number of tourism regions. Many tourism-relevant data, such as points of interest, ski resorts, user reviews, restaurants, job descriptions, accommodation equipment including dynamic content such as prices or availability is available on websites, but are only used sporadically by RTOs. Despite the comparable contents on the websites of RTO and a common objective to achieve the highest possible online visibility and better presentation in the search results and thus a higher booking and attendance rate, the usage scenarios of Semantic Web differ in Austrian tourism regions.

The findings of this study are based on a secondary source. This implies that the number of items of investigation was limited from the start. It has not been investigated whether the sites selected for this analysis were included in the original 3.2 billion site list. In addition, only the websites with a maximum of four website navigation levels were included in the original data set. The original record may also exclude websites that prohibit the browsing of their contents by the unknown web crawlers, which was also not checked during this analysis. Furthermore, the structured data was extracted

from the dataset for November 2017 at a single point in time, making it impossible, for example, to check some records in real time. An interesting research approach for the future would be to repeat the same study at a periodic interval to see if the use of Semantic Web technology has changed over time.

Another limitation of this study is the fact that several errors in the semantic annotations on the websites were found when preparing the source data for analysis. Such mistakes not only complicate data analysis but also may fail the very purpose of structured data. Since systematically error detection was not subject of this work, these may bias the analysis results through wrong classification or incorrect detection of semantic markup. Future research should focus more on error analysis in semantic annotations and how these errors could be avoided (e.g. through semantic annotation tools).

The analysis results may have further been influenced by the non-differentiation of language variants of a website. Thus, tourist regions with a large number of indexed pages on search engines, representing many touristic objects in multiple languages show better results in this analysis. In addition, the proportion of structured data that was used only on the subdomains of the websites of RTO has not been determined. It is thus possible that a whole tourist region shows better results, even though semantic annotations were only made on a few subdomains. Thus, an international comparison that copes with different languages and/or subdomains would be of interest. This would allow identifying best practices and recommended actions specifically for the tourism organizations in a certain country.

Even though tourism-specific semantic markup is not widely used in Austrian RTO websites, it can be expected that with the increasing spread of intelligent web applications and services, more and more content owners will deal with this subject. A better visibility of the services and offers of the touristic region through semantic annotations helps in the dissolution of dependence on international online intermediaries and should therefore be more widespread in the websites of Austrian tourism organizations.

References

1. Statistics Austria: Arrivals, Overnight Stays. http://www.statistik.at/web_en/statistics/Economy/tourism/accommodation/arrivals_overnight_stays/index.html. Accessed 05 Feb 2019
2. Statistics Austria: A tourism satellite account for Austria. http://www.statistik.at/web_en/statistics/Economy/tourism/tourism_satellite_accounts/value_added/index.html. Accessed 05 Feb 2019
3. UNWTO: World Tourism Barometer **16**(1), 1–26 (2018)
4. Franch, M., Martini, U., Inverardi, P.L.N., Buffa, F.: The role of the regional tourist boards in the destination marketing policies. The case of the dolomites. *Int. Rev. Public Nonprofit Mark.* **1**, 113–124 (2004)

5. Fensel, A., Kärle, E., Toma, I.: TourPack: packaging and disseminating touristic services with linked data and semantics. In: Hölldobler, S., Liang, Y. (eds.) *Proceedings of the 1st International Workshop on Semantic Technologies (IWOST)*, pp. 43–54. CEUR-WS.org (2015)
6. Stavrakantonakis, I., Toma, I., Fensel, A., Fensel, D.: Hotel websites, Web 2.0, Web 3.0 and online direct marketing: the case of Austria. In: Xiang, Z., Tussyadiah, I. (eds.) *Information and Communication Technologies in Tourism 2014*, pp. 665–677. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-03973-2_48
7. Toma, I., Stanciu, C., Fensel, A., Stavrakantonakis, I., Fensel, D.: Improving the online visibility of touristic service providers by using semantic annotations. In: Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (eds.) *ESWC 2014. LNCS*, vol. 8798, pp. 259–262. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11955-7_31
8. Kärle, E., Fensel, D.: Annotation based automatic action processing. In: Nikitina, N., Song, D., Fokoue, A., Haase, P. (eds.) *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks* (2017)
9. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. *Int. J. Semant. Web Inf. Syst.* **5**, 1–22 (2009)
10. Kärle, E., Fensel, A., Toma, I., Fensel, D.: Why are there more hotels in Tyrol than in Austria? Analyzing Schema.org usage in the hotel domain. In: Inversini, A., Schegg, R. (eds.) *Information and Communication Technologies in Tourism 2016*, pp. 99–112. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28231-2_8
11. Buhalis, D., Law, R.: Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tour. Manag.* **29**, 609–623 (2008)
12. Navio-Marco, J., Ruiz-Gómez, L.M., Sevilla-Sevilla, C.: Progress in information technology and tourism management: 30 years on and 20 years after the internet - Revisiting Buhalis & Law’s landmark study about eTourism. *Tour. Manag.* **69**, 460–470 (2018)
13. Vandenbussche, P.-Y., Atemezing, G.A., Poveda, M., Vatan, B.: Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web* **8**, 437–452 (2017)
14. Linked Open Vocabularies (LOV). <https://lov.linkeddata.es/dataset/lov>. Accessed 18 Feb 2019
15. Kärle, E., Simsek, U., Akbar, Z., Hepp, M., Fensel, D.: Extending the Schema.org vocabulary for more expressive accommodation annotations. In: Schegg, R., Stangl, B. (eds.) *Information and Communication Technologies in Tourism 2017*, pp. 31–41. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-51168-9_3
16. Soualah-Alila, F., Faucher, C., Bertrand, F., Coustaty, M., Doucet, A.: Applying semantic web technologies for improving the visibility of tourism data. In: Balog, K., Dalton, J., Doucet, A., Ibrahim, Y. (eds.) *Proceedings of the Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval - ESAIR 2015*, pp. 5–10. ACM Press, New York (2015)
17. Jakkilinki, R., Sharda, N.: A framework for ontology-based tourism application generator. In: Pease, W., Rowe, M., Cooper, M. (eds.) *Information and Communication Technologies in Support of the Tourism Industry*, pp. 26–49. Idea Group Pub, Hershey (2007)

18. Fodor, O., Werthner, H.: Harmonise: a step toward an interoperable e-tourism marketplace. *Int. J. Electron. Commer.* **9**, 11–39 (2005)
19. Ou, S., Pekar, V., Orasan, C., Spurk, C., Negri, M.: Development and alignment of a domain-specific ontology for question answering. In: Proceedings of the 6th Edition of the Language Resources and Evaluation Conference, LREC 2008 (2008)
20. Barta, R., Feilmayr, C., Pröll, B., Grün, C., Werthner, H.: Covering the semantic space of tourism. In: Gómez-Pérez, J.M. (ed.) Proceedings of the 1st Workshop on Context, Information and Ontologies, CIAO 2009, Heraklion, Greece, 1 June 2009, pp. 1–8. ACM Press, New York (2009)
21. Hepp, M.: Accommodation Ontology Language Reference. <http://purl.org/acco/ns>. Accessed 18 Feb 2019
22. Gazzè, D., Lo Duca, A., Marchetti, A., Tesconi, M.: An overview of the tourpedia linked dataset with a focus on relations discovery among places. In: Hellmann, S., Parreira, J.X., Polleres, A. (eds.) SEMANTiCS Vienna 2015. Proceedings of the 11th International Conference on Semantic Systems: 16th–17th of September 2015, Vienna, Austria, pp. 157–160. The Association for Computing Machinery, New York (2015)
23. Meusel, R., Petrovski, P., Bizer, C.: The WebDataCommons microdata, RDFa and microformat dataset series. In: Mika, P., et al. (eds.) ISWC 2014. LNCS, vol. 8796, pp. 277–292. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11964-9_18
24. Meusel, R., Bizer, C., Paulheim, H.: A web-scale study of the adoption and evolution of the schema.org vocabulary over time. In: Akerkar, R., Dikaiakos, M., Achilleos, A., Omitola, T. (eds.) Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, WIMS 2015. ACM Press, New York (2015)
25. Cao, K., Yang, Z.: A study of e-commerce adoption by tourism websites in China. *J. Destin. Mark. Manag.* **5**, 283–289 (2016)
26. del Carmen Calatrava Moreno, M., Hörhager, G., Schuster, R., Werthner, H.: Strategic E-Tourism alternatives for destinations. In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, pp. 405–417. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14343-9_30
27. Luna-Nevarez, C., Hyman, M.R.: Common practices in destination website design. *J. Destin. Mark. Manag.* **1**, 94–106 (2012)
28. Meusel, R., Paulheim, H.: Heuristics for fixing common errors in deployed *schema.org* microdata. In: Gandon, F., Sabou, M., Sack, H., d’Amato, C., Cudré-Mauroux, P., Zimmermann, A. (eds.) ESWC 2015. LNCS, vol. 9088, pp. 152–168. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18818-8_10
29. Şimşek, U., Kärle, E., Holzknicht, O., Fensel, D.: Domain specific semantic validation of schema.org annotations. In: Petrenko, Alexander K., Voronkov, A. (eds.) PSI 2017. LNCS, vol. 10742, pp. 417–429. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74313-4_31
30. Kärle, E., Şimşek, U., Fensel, D.: semantify.it, a platform for creation, publication and distribution of semantic annotations. In: Homenda, W., Roman, D. (eds.) The 11th International Conference on Advances in Semantic Processing (SEMAPRO), pp. 22–30 (2017)
31. Hepp, M., Siorpaes, K., Bachlechner, D.: Towards the semantic web in e-tourism: can annotation do the trick? In: ECIS 2006 Proceedings (2006)

32. Akbar, Z., Kärle, E., Panasiuk, O., Şimşek, U., Toma, I., Fensel, D.: Complete Semantics to empower Touristic Service Providers. In: Panetto, H., et al. (eds.) OTM 2017, vol. 10574, pp. 353–370. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-69459-7_24
33. Zanker, M., Fuchs, M., Seebacher, A., Jessenitschnig, M., Stromberger, M.: An automated approach for deriving semantic annotations of tourism products based on geospatial information. In: Höpken, W., Gretzel, U., Law, R. (eds.) Information and Communication Technologies in Tourism, pp. 211–221. Springer, Vienna (2009). https://doi.org/10.1007/978-3-211-93971-0_18
34. Krajasits, C., Andel, A., Wach, I.: Stellenwert der Gemeinden für den österreichischen Tourismus. https://www.oir.at/files/download/projekte/Raumplanung/Tourismusgemeinden_EB_Sep08.pdf. Accessed 07 Feb 2019

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

