





Article

OntoTouTra: Tourist Traceability Ontology Based on Big Data Analytics

Juan Francisco Mendoza-Moreno ^{1,*}, Luz Santamaria-Granados ¹, Anabel Fraga Vázquez ² and Gustavo Ramirez-Gonzalez ³

¹ GIDINT, Faculty of Systems Engineering, Universidad Santo Tomás Seccional Tunja, Calle 19 No. 11-64, Tunja 150001, Colombia; luz.santamaria@usantoto.edu.co

² Knowledge Reusing Group, Computer Science, Universidad Carlos III de Madrid, Av. de la Universidad, 30, 28911 Madrid, Spain; afraga@inf.uc3m.es

³ GIT, Telematics Department, Universidad del Cauca, Calle 5, No. 4-70, Popayán 190002, Colombia; gramirez@unicauca.edu.co

* Correspondence: juan.mendoza@usantoto.edu.co; Tel.: +57-(2)-820-9900

Abstract: Tourist traceability is the analysis of the set of actions, procedures, and technical measures that allows us to identify and record the space–time causality of the tourist’s touring, from the beginning to the end of the chain of the tourist product. Besides, the traceability of tourists has implications for infrastructure, transport, products, marketing, the commercial viability of the industry, and the management of the destination’s social, environmental, and cultural impact. To this end, a tourist traceability system requires a knowledge base for processing elements, such as functions, objects, events, and logical connectors among them. A knowledge base provides us with information on the preparation, planning, and implementation or operation stages. In this regard, unifying tourism terminology in a traceability system is a challenge because we need a central repository that promotes standards for tourists and suppliers in forming a formal body of knowledge representation. Some studies are related to the construction of ontologies in tourism, but none focus on tourist traceability systems. For the above, we propose OntoTouTra, an ontology that uses formal specifications to represent knowledge of tourist traceability systems. This paper outlines the development of the OntoTouTra ontology and how we gathered and processed data from ubiquitous computing using Big Data analysis techniques.

Keywords: tourist traceability; ontology; Big Data; analytics; ubiquitous computing

check for
updates

Citation: Mendoza-Moreno, J.F.; Santamaria-Granados, L.; Fraga Vázquez, A.; Ramirez-Gonzalez, G. OntoTouTra: Tourist Traceability Ontology Based on Big Data Analytics. *Appl. Sci.* **2021**, *11*, 11061. <https://doi.org/10.3390/app112211061>

Academic Editors: Jesualdo Tomás Fernández Breis

Received: 27 September 2021
Accepted: 10 November 2021
Published: 22 November 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The relationship between the concept of traceability and the tourist contributes to the improvement of the methodological approaches used in studies because it provides us with the precision and validity of the data obtained, especially from ubiquitous environments [1]. Traceability constitutes an advance for the collection of tourist mobility data in spatial–temporal relationships. Traditionally, in the fields of production, logistics, and software, traceability has been considered as the set of actions, metrics, and technical procedures to identify and record each product from the beginning to the end of the supply chain [2]. Furthermore, the ISO defines the traceability concept “as the ability to trace the history, application, or location of that which is under consideration.” [3]. Furthermore, the GS1 defines tracing as “the ability to identify the origin, attributes, or history of a particular traceable item” and tracking as “the ability to follow the path of a traceable item” [4].

In this sense, through a TTS, the DMO can identify the routes of the tourists and the degree of interest that the attractions of the destination arouse in them. Furthermore, TTS can use sociodemographic metrics and statistics reports to identify tourist profiles, to prepare and adapt both the tourist destination and the tourism management system. Hence, with the accelerated technological advance that characterizes ubiquitous computing, now

DMOs have at their disposal various data sources. These sources provide input data for the TTS, such as social networks, cloud platforms, the web, the IoT, traditional databases, public or private datasets, and linked data, among other data sources.

On the other hand, these data sources typically are extensive volume datasets and reach high speed (in real time or almost in real time). Furthermore, variety is another characteristic of these data (some have a format; the vast majority do not). Big Data can process and store this type of data and constitute a knowledge base through ontological systems. In this way, the DMO can make decisions based on the information processed.

Currently, in most cases, the DMO makes decisions based on paper surveys applied to some tourists. Furthermore, government reports and those of the tourism sector actors serve as data for this decision-making process. These strategies have drawbacks, such as the subjectivity and predisposition of tourists to answer surveys. Many of them prefer not to answer them for time or data privacy reasons, and government reports are generated in extended periods, and in some cases, they arrive late. For this reason, the research gap of this study arises, which takes advantage of data from ubiquitous sources to provide information related to the traceability of tourists to a given destination. In this way, with the processing of these characteristic Big Data, precisely due to the volume, velocity, and variety, to constitute a knowledge base, the research question arises: How can we develop a tourist traceability ontology based on obtaining data and ubiquitous data processing, using Big Data analytics techniques?

It is worth mentioning that the purpose of this study is to constitute an ontology based on data previously generated in a massive way, not on data from tourists in particular. Initially, we considered the data from three types of ubiquitous sources: reviews of tourists in OTAs, data from sensors located in the POIs of the destination, and data from tourist guide applications installed on the tourist's mobile devices, which have prior permission for further processing. A tourist traceability ontology allows the DMO to make decisions regarding the management of the destination according to the flow and track of tourists, determine their preferred POIs, intelligently dispose of the infrastructure for adequate attention, and foresee improvements in services, as well as design tourist experiences according to the interests of the tourist in a space–time causality.

OntoTouTra is an ontology that explains the structure of knowledge, whose domain is the tourist traceability system, based on data collected from ubiquitous systems. OntoTouTra shares this knowledge through the conceptual design of this domain, enabling the reuse of knowledge. This paper shows the development of the OntoTouTra ontology. The ontology input data are from pervasive data sources. OntoTouTra is useful for making decisions on the destination, and its ontological goal [5] is the integration of homogenous or heterogeneous data sources and search engines and building knowledge systems.

The paper is organized as follows: Section 2 reviews the development of tourism ontologies related to this research. In Section 3, we explain the ontology structure. Then, in Section 4, we describe the methodology of the Big Data analytics used to build the ontology. Section 5 provides the details for the implementation of the ontology with the chosen tourist destination. In Section 6, we show the results obtained from the experiment. Furthermore, Section 6 shows the knowledge base for the tourist traceability system, and Section 7 describes the data treatment for OntoTouTra. Then, Section 8 highlights the conclusions of the study and future work. Finally, this paper has Supplementary Material, which is an ontology implementation document. There are lists and figures that illustrate the process of creating OntoTouTra and the different possibilities of queries according to the requirements for decision-making in the management of a tourist destination in the domain of the TTS.

2. Related Work

Some research on the semantic representation of the tourism domain uses information gathered from tourism websites for different applications. Xiang et al. [6] concluded that tourism websites can incorporate tools (such as reviews, tagging, and excavation) to

allow travelers to interact directly with these sites. This way, the knowledge of travelers' perceptions and experiences can be collected and learned. Therefore, these tools offer promising avenues for tourist destination specialists to better understand and interact with potential visitors. Hence, the ontology is "the language of tourism" between the traveler and the industry. OTAs have communication channels with tourists to interact between them and the operators. Mainly, these channels are based on reviews that tourists give about their experience; in general, OTAs tag these reviews.

Concerning the knowledge domain of tourism, research such as that of Tribe and Liburd [7] "reconceptualizes" its system, taking into account three cores: disciplinary knowledge, problem-centered knowledge, and value-based knowledge. The domain of the TTS refers to the disciplinary knowledge of the ontology of this research. It denotes the importance of understanding that tourism is a multidisciplinary field and an extradisciplinary one and considers the person, position, ideology, government, and global capital as elements of expertise. The "problem-centered knowledge" lies in the fact that DMOs need to have a knowledge base for decision-making, especially statistical information obtained from the traceability of the tourist at the destination. The decision-making by the DMO enables the improvement of the destination infrastructure and the feedback of the tourism management system; in this way, we obtain value-based knowledge.

Mouhim et al. [8] highlighted the importance of KM in tourism: share knowledge, facilitate the development of new products and services, develop the ability to learn, acquire tacit knowledge to transform it into explicit knowledge, satisfy customers, and exploit the market. Based on [9], these researchers analyzed existing ontologies such as the Harmonize Ontology [10], for the exchange of data between organizations; the Mondeca Ontology [11] for profiling tourist and cultural objects, tourist packages, and multimedia content for tourism; and the OnTour project [12], which describes the domain of tourism focused on accommodation and activities. Seeing that none of these ontologies met the particular needs of their destination city, the researchers created their ontology (the Moroccan Tourism Ontology), taking advantage of the thesaurus of UNESCO and the UNWTO. Its ontology has the main classes: accommodation, transportation, attractions, activities, services, restaurants, and cultural heritage. As a study preceding their OnTourism project, Prantner et al. [13] analyzed, in addition to the ontologies above, the OTA specification, the Tourism Ontology of the University of Karlsruhe, and the traveling ontologies EON and TAGA. They also reviewed the main ontology management tools for the domain of tourism, identifying the following: DIP Ontology Management Suite, WSMT, WebOnto, and Ontolingua [14]. They complement the previous state-of-the-art because they feature a summary of ontologies in the travel industry, adding to the Comprehensive Ontology for the Tourism Industry, the LA_DMS project for destinations, the SWAP project, the Tiscover platform, and the Hi-Touch project, for the domain of intra-European sustainable tourism. The ontology proposed in this document has as its domain the tourist traceability system in a specific destination. In contrast to the ontologies described above, we used Big Data analysis for the building of OntoTouTra. We collected data from ubiquitous computer sources, especially from social networks.

Subsequently, in the process of building a domain ontology for African tourism areas, Zhao et al. [15] reviewed new ontologies such as the e-tourism ontology, Tourism ProtegeEsportOWL, and the botanical ontology of the National Knowledge Infrastructure of the Chinese Academy of Sciences. In this way, they proposed a method for the construction of ontologies in seven steps: determine the field and the scope; examine existing ontologies; summarize essential concepts; define the classes and their hierarchy; define the attributes; define the properties; and finally, establish the individuals. We analyzed the Big Data analytical methodology proposed by Erl et al. [16], in addition to contemplating the steps of the previous methods, as being ideal for the collection and processing of large volumes of data at high transfer rates.

To unify the tourism terminology, we need a central authority that promotes standards for tourists and suppliers to understand tourism-related ontologies. Huang and Bian [17] recognized the UNWTO's effort in defining the thesaurus about tourism and leisure activities, but believed that it is not enough due to the complex character of tourist data. They proposed their research to integrate both types of ontologies through the formal concept analysis and Bayesian approaches. These approaches are mathematical tools for data analysis, knowledge representation, and information management, using triples with binary relations among the concepts.

More recent studies, such as Valls et al. [18], entrusted their research to word ontologies, such as WordNet [19], applying clustering based on ontologies, determining the motivations of tourists when visiting a destination. OnTraNetBD [20] also used WordNet for mapping the key concepts to build the ontology using the Domain, Entity Classes, Relations, and Attributes (DERA) methodology [21] in six phases: identify atomic concepts, analysis, synthesis, standardization, ordering, and formalization. From WordNet and Wikipedia were derived Yet Another Great Ontology (YAGO) [22], which uses a logical model, capable of representing n-ary relations maintaining compatibility with the RDFS. In this sense, Reference [23] developed a system that supports different types of document formats, including the essential structures of textual documents and native forms of the web. In the paper, the authors compared the results of the semantic annotation approach with other popular methods (Armadillo, CERNO, CREAM, EVONTO, GoNTogle, KIM, MnM, Onto-Mat, and S-CREAM). Ontologies based on words use relations between elements, for instance, Llorens et al. [24] called the words "terms" and established the relationships among the terms as the entity relationship model of the UML diagrams in software engineering.

The tourism sector has highlighted the need to develop personalized applications using knowledge bases. Currently, researchers focus their interest on the development of applications based on ontologies. Such is the case of the scientometric review that we preliminarily carried out on the frameworks of tourist recommendation systems [25] that use heterogeneous data sources extracted from wearable devices, the IoT, social networks, and ontologies. A specific application we found is the TRSO [26] recommendation system for tourists to know the attractions they can see and the activities they can do. The recommender system uses collaborative filtering techniques based on information from attraction ontologies. Investigations such as SocioOntoProcess [27] draw from social networks to build ontologies and take advantage of user interactions to develop the models, in this case for consulting a consensual vocabulary. The ontology's construction is collaborative through web tools, such as wikis.

SigTur/E-Destination [28] is a project that, from the knowledge management point of view, through a specific domain ontology, provides information on activities and guides aimed at the user and for employees. The system considers as much information as possible (demography, spatial, travel, motives, user stereotypes) to make the recommendations. We were also motivated to gather data from social networks, especially from OTAs or eWOMs, because they tagged tourist reviews. Some OTAs offer an API to consult these reviews, but it is necessary to develop tools that can collect those public reviews for others. For this purpose, we created a web scraping tool.

From the perspective of the software industry, in particular the reuse of information, arose the RSHP meta-model [24]; the authors looked for a general model capable of representing the information of software artifacts, without dependence on their internal structure. They found that the data of all the artifacts form a representation of a particular domain. The authors concluded that the field could be created automatically by indexing the artifacts through a fundamental and simple idea: "the information is related facts." Therefore, the central element of an artifact is the relationship. The semantics of the RSHP qualifies the existing relationship and its type; its components are artifact, term, relationship, information element, and property.

During the last decade, Shoval and Ahas [29] reviewed the literature on the use of tracking technologies for tourism, finding forty-five articles (40% of the articles were published in the three leading tourism journals). This review found that tracking data occur in three generations: The first generation deals with methodological research and analyzes the potential of tracking data. The second generation is related to spatial and temporal data. The third generation is interested in new data sources. The researchers concluded that the movement of tourists has implications for infrastructure, transport, products, marketing, the commercial viability of the industry, and the management of the social, environmental, and cultural impact of the destination. They also detected the current research gaps in this area: a large amount of data for processing, personal data, and tourist data protection. Using new techniques is necessary to know the tourist traceability since some theorists think that the tourist can change the activity or behavior when being followed or studied.

Girardin et al. [30] proposed a challenge for social science research since large volumes of data from ubiquitous sources are available. With these data, we can understand the dynamics of the population and customize the services, among other essential activities for tourism management. They named the tourist tracks “digital footprints”, which are of two types: active and passive. The passive traces are data left with the interaction of infrastructure, and the active traces are the location data exposed by the users, especially in social networks. They worked with Flickr data (actives) and the call records of a telephone company (passives). The data used in Flickr are explicitly public data by the user. They carried out the process and the visualization of the large volumes of data through geo-visualization. Concerning data privacy, the authors handled the number of users instead of individual data. For this research, the expression “digital footprints” is similar to the data sources of ubiquitous computing, which are the input of the traceability system.

Mariani and Borghi [31] conducted a review of research literature on hospitality and tourism with Big Data and Business Intelligence to identify future research and development gaps. They found that the research that applied analytical techniques is limited in scope and methodologies. Besides, conceptual frameworks are missing to identify critical business problems that link Business Intelligence and Big Data to tourism management. They evidenced epistemological dilemmas for the development of knowledge theories conducted using Big Data. They concluded with their study that further research on tourism should be stimulated and systematized by leveraging Big Data and Business Intelligence and providing information bases aimed at companies and stakeholders in tourism.

As a synthesis of this review of the related work, Table 1 depicts the highlighted ontologies and their respective objectives.

In Table 1, we see that all ontologies meet a particular objective, which is why their domain of knowledge is well defined. We show that none of the ontologies listed in this table have tourist traceability as their domain.

Chantre et al. [1] established two thematic cores of the movement of tourists and the tracking methodologies in the relationship of traceability and the tourist. In this sense, they considered tourist traceability as the set of actions, measures, and technical procedures to identify and record the activity of tourists in a given destination. For the above, to keep this record, it is necessary to build a spatiotemporal causality. Through a tourist traceability system, we gather information on the activities of interest to tourists, the most frequented POIs, the timing of visits, tourist satisfaction with their experience, visitor profiling, and a portfolio of tourist experiences, among others. In turn, a TTS allows decision-making by the DMO, establishing KPIs that determine the level of service offered to improve destination management. With the above considerations, it is essential to have a knowledge base of the TTS domain, with updated, accessible, actionable, and reliable data.

Table 1. Tourism domain ontologies found in the literature review.

Ontology	Year	Purpose	TTS Concepts Covered?
Architectural ontology [32]	2018	e-tourism resources	No. It has an architectural domain.
OnTraNetBD [20]	2017	Uses WorNet for mapping key concepts	No. The ontology establishes the formal relationship between tourist attractions and other travel elements, but not the space–time causality of the tourist.
Ontology-Based Tourism Recommendation System [33]	2017	Travel ontology	Partially. It defines a travel recommendation system based on ontologies, but does not analyze tourists' routes in the destination.
Ontology-Based Human–Computer Cloud [34]	2017	Building ad hoc decision support services	No. It describes various decision support scenarios in tourism in general, but not specifically for the TTS.
Dwipa Ontology III [35]	2017	Cultural parks, artists, and monuments	No. It is limited to POIs.
TRSO [26]	2016	Recommender system for tourists	Partially. It determines the relationship of tourists with the context to suggest tourist information.
SigTur/E-Destination [36]	2011	Activities and guides	No. It provides a catalog of destination resources to offer personalized information to tourists.
Mondeca [13]	2011	Profiling tourist and cultural objects	Partially. Mondeca has a large number of concepts on tourism, but it is not freely available.
Moroccan Tourism [8]	2011	Ontology of this destination city	No. It is limited to presenting the importance of the knowledge domain in tourism.
University of Karlsruhe [13]	2007	OnTourism project for evaluating the Semantic Web	No. They analyzed seven tourism ontologies and five management tools to create ontologies.
OnTour project [12]	2006	Accommodation and activities	No. It focuses to e-tourism.
Harmonize Ontology [10]	2004	Exchange data between organizations	No. It is aimed at developing an interoperability platform for SMEs in the tourism sector.

This study took advantage of data from ubiquitous sources, especially from OTAs, because these satisfy the above requirements, especially tourist reviews. Furthermore, these allow identifying, among others, data on spatiality, temporality, satisfaction, feelings, preferences, and experiences. The analysis of these data is boosted through link data, for instance, with georeferenced data from tourist reviews, we reach more location levels, establishing a relationship between the review location and the hotel, destination, POI, or service reviewed. We move up the geographical level, passing through the state or region and reaching a particular country. Linking data with GeoNames provides complementary geographic information, which we did not obtain directly from the ubiquitous data source. Similarly, complementary temporal information is collected from linking data with the Time Ontology.

The GeoNames ontology [37] allows adding semantic data to the World Wide Web. It has more than 11 million toponyms with a single URL (RDF web service). The ontology of GeoNames is available in OWL as a database dump and also as open data linked in RDF [38]. Geographic levels in GeoNames [39] vary according to the country, for example, Germany has six levels, France five, and Colombia four. Therefore, it was necessary to resort to national data providers; for the Colombian case, the National Administrative Department of Statistics (DANE) provides the DIVIPOLA system [40]. Thus, we can provide more data about the location of a person, hotel, or tourist attraction (POI).

The other aspect of the spatiotemporal relationship of tourist traceability is based on temporal concepts; the OntoTouTra data link is the Time Ontology [41]. We took advantage of the vocabulary from this ontology to express the facts of relations between instants and intervals. We can establish temporal reference systems (time: *DateTimeDescription*), position in time (time: *TemporalPosition*), intervals (time: *DateTimeInterval*), and duration (time: *duration time: DurationDescription*).

OntoTouTra does not have a data link with any tourism management ontology. However, for its construction, open data repositories were taken into account by the International Open Data Charter [42], for instance, we used Colombia's Open Data [43] and SITUR [44].

3. The Ontology: OntoTouTra

3.1. Tourist Traceability System

A TTS allows recording a tourist's history, application, or location while touring the destination's different POIs. From this tour through the POIs, we gather various types of data: location, time, permanence, preference, indifference, assessment, suggestions, and recommendations, among others. The sources of data can be diverse: traditional sources such as surveys, suggestion boxes, government and tourism industry reports, hotel occupancy, transport sector reports on travelers, among others; even ubiquitous computing sources, such as sensors, the IoT, mobile devices, social networks, apps, the web, and software applications. For ubiquitous sources, data have these features: velocity, volume, and variety. Based on the data, the DMO can make decisions concerning the improvement of the destination's infrastructure, the tourist management system, the promotion, and the marketing of the tourist destination, in search of satisfying the tourist's expectations, needs, and contentment. Tourism is a dynamic sector; it depends on the context and profile of the tourist. Decision-making tends to be a complicated process, and it is a real challenge to obtain and process data and generate information and knowledge.

In recent studies, researchers in the tourism domain have aroused interest in investigating patterns of mass movements of tourists as they are touring a destination. However, gaps have been identified concerning this movement since the investigations tend to be a holistic analysis. We are interested in studying the tracking, the trace, the routes, and the tourist's behavior in a space–time causality. This approach is similar to traceability investigations in supply chains, software development, healthcare, and security. Traceability is the ability to trace something and verify an item's history, location, or application context through

recorded and documented identification. Therefore, tourist traceability provides valuable elements for decision-making in managing a tourist destination [1]. A TTS can provide information to answer questions. Some of these questions are:

- **POI:** What are the busiest POIs? What type of visitors frequent them? In what time slot are they visited? Where do the tourists come from? Later, where do they go? What activities do they mostly do? What tourist experiences are enjoyed?
- **Seasonality:** What is the behavior of seasonality in the destination? What activities are carried out due to seasonality? What services do they consume due to seasonality? What is the offer of tourist experiences?
- **Suppliers:** What is the level of satisfaction with the services provided? What are the needs to satisfy the demand?
- **Stakeholders:** How do stakeholders interact at the beginning, during, and at end of the visit to the tourist destination? What suggestions do tourists have regarding this service chain?

3.2. OntoTouTra Analysis

The interaction of the tourist with the destination POIs is the subject matter of the ontology. The DMO needs to know what experiences the tourist had and their degree of satisfaction to facilitate the decision-making process to improve the destination. We collected data from sources of pervasive computing (mainly social networks) and government and tourism sector sources.

The domain of this ontology has as its main classes: the DMO, tourism experiences, tourist attractions (POIs), and destinations. We established the relations within the tourist domain:

- DMOs provide the service that the tourist consumes;
- The tourists live the experiences in the destination;
- The tourist attractions are the push factor and motivator for the tourist;
- The destination is the geographical location where tourist traceability happens.

These four relationships were the starting point of the ontology design; we designed the use case diagram (see Figure 1) and created the primary classes of the ontology mentioned above. From these classes, we generated the subclasses, properties, and relationships between classes.

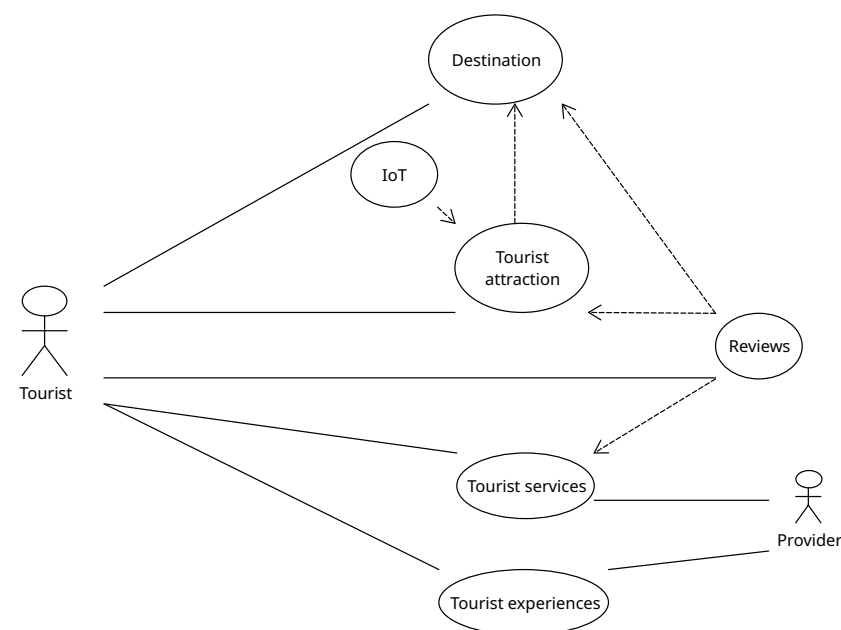


Figure 1. Tourist traceability system: use case.

The top-down design of the ontology provided the hierarchical order of the terms, starting from the root domain, that is to say, the tourist traceability (root node), and distributed by the general classes until arriving at the specific terms. Identifying the terms from sources of authorities on tourism and other similar ontologies and with the DMO's expertise thus ensures the formulation and definitions of the ontology's taxonomic hierarchy. The process of the revision of iterative versions is necessary to guarantee the consistency of the definition and the scientific, logical, and philosophical rigor of the terms (see Figure 2).

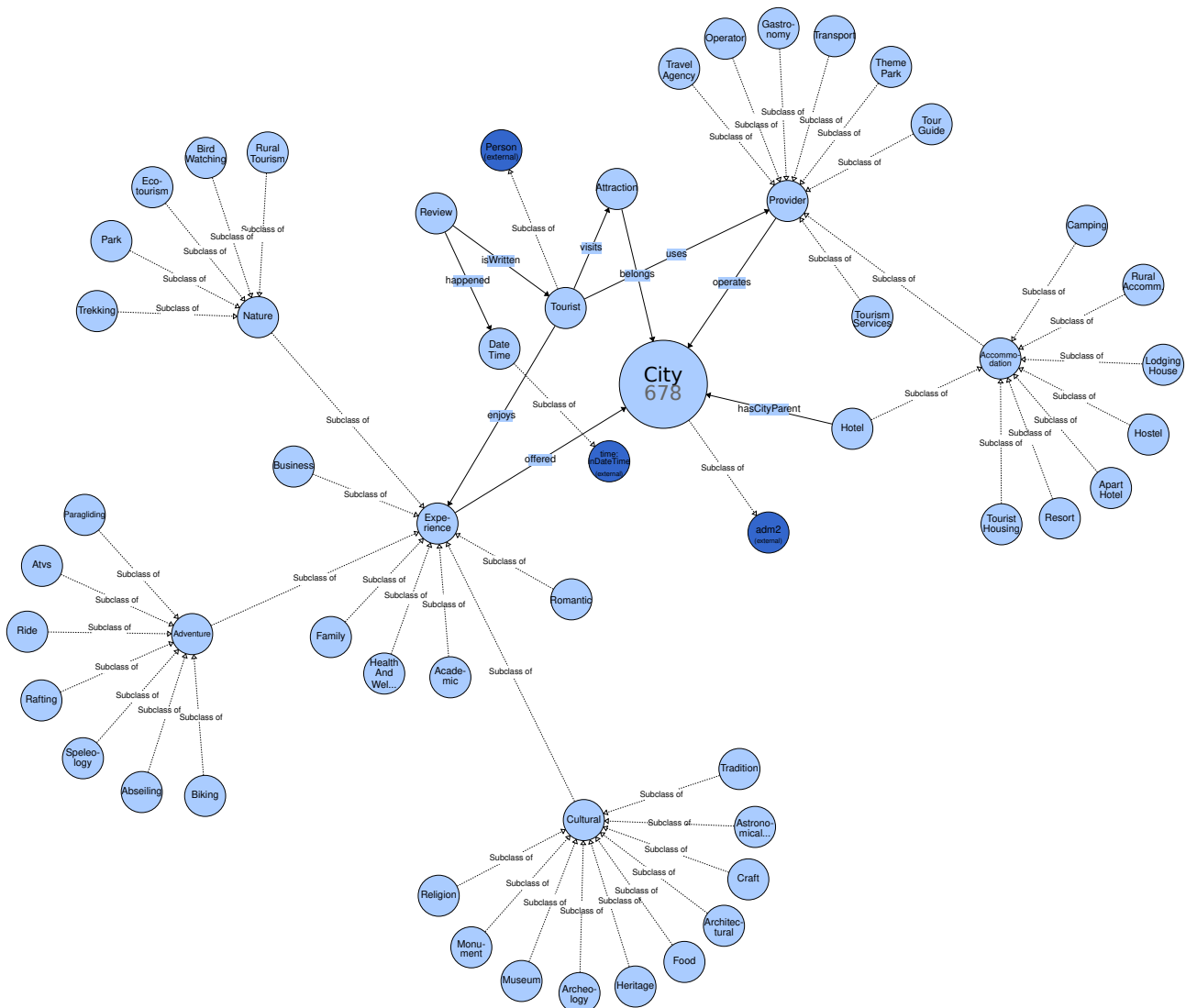


Figure 2. Snippet of the image of the upper levels of OntoTouTra (using WebVOWL [45]).

3.3. Development of the Ontology on the Domain of Tourism Traceability

We focused on using a method for the ontology's construction, such as METHONTOLOGY [46], a methodology that allows building ontologies from scratch and has been tested in different knowledge domains. Using this methodology, we took advantage of the Big Data analytics lifecycle model [16] to obtain, process, classify, and visualize data from ubiquitous computing sources. This ontology, called OntoTouTra, has as its principal purpose to provide a knowledge base to handle problems of semantic aspects to support the implementation of a TTS.

METHONTOLOGY [46] uses an iterative approach to tailor the ontology to refine the TTS domain model. In this way, we moved from the level of knowledge (conceptual model)

to the level of implementation (logical or computational model), looking for the ontology to be readable by machines. For the construction of the conceptual model (see Section 3.4), we began with the identification of the purpose and scope of the ontology, as follows:

3.3.1. Specification

The domain of the ontology is the TTS with four main branches: Provider, Tourist Experience, Destination, and Tourist (see Figures 2 and 3). Understanding these branches avoided any inconsistencies between the classes and the ontology. In addition, these branches responded to the requirements of tourism traceability: Where are the tourists (Destination)? What do tourists do (Tourist Experiences)? Who offers the experiences (Provider)? From these branches were derived the classes that make up the TTS domain. The POIs and the tourist reviews are important because they implemented the space–time relationship to answer questions of the domain: When and where does the tourist consume the experience, or what is the tourist’s opinion of the experience?

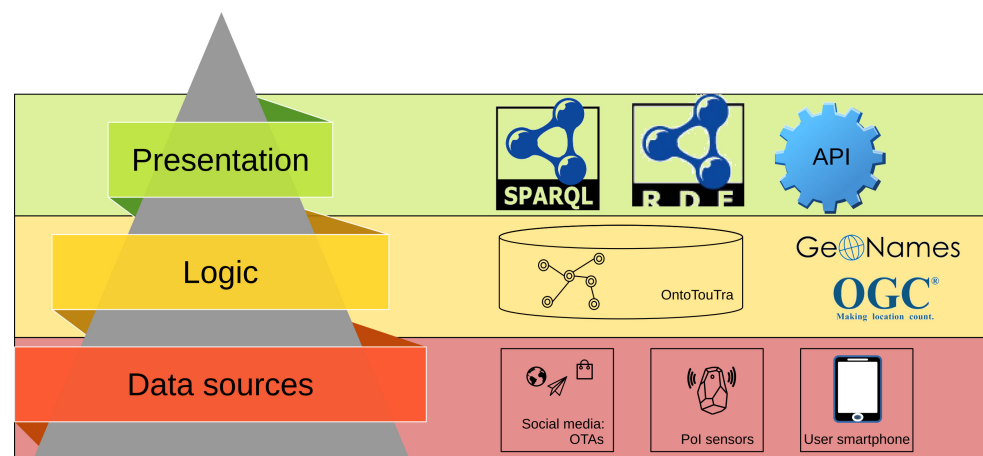


Figure 3. OntoTouTra architecture.

3.3.2. Conceptualization

We considered two types of data sources for knowledge acquisition (see Table 2). The first source corresponds to expert organizations in the tourism domain. We analyzed the management documentation, policies, guidelines, and reports to define the ontology domain’s branches, classes, subclasses, relationships, properties, and scope. The second data source is ubiquitous computing. In this phase, we refined the ontology, iterating between the specification and the conceptualization. The data were gathered with web scraping from the OTA. Due to its Big Data characteristics, we applied the appropriate methods for these environments, such as data mining, text mining, and the MapReduce technology. With the vocabulary obtained (see Section 4.2.3), mainly from tourist reviews, we refined the previously defined concepts with the first data source type.

Subsequently, we built the TTS glossary, identifying the concepts and ensuring each term was described with synonyms and acronyms. We also checked the terms that referred to the same concept (related terms). Each term has a simple description within the ontology. Through the relationships between classes, we avoided any ambiguity of concepts, for instance: destination-city-municipality, point of interest-POI-attraction, tourist-visitor-reviewer, and provider-supplier. A fragment of the TTS glossary is depicted in Table 3.

We implemented the top-down approach (see Section 3.4), starting with a general level until the level of the details. By identifying the classes and their relationships, we defined the taxonomy and hierarchy of the ontology. According to Kumara et al. [47], a *hierarchy* is defined as $H = (N, E)$, which is a simple directed graph, where N is the nodes and a set of edges $(n_p, n_c) \in E \subseteq N \times N$. The address of an edge (n_p, n_c) is defined from the parent node n_p to the child node n_c (SubClass Of).

Table 2. Data sources of the individuals of the main classes of OntoTouTra.

Ontology Main Class	Data Source (Individuals)	Linked Data	Data Sources Used in This Research
Tourist	social networks: OTA, eWOM	foaf	[48–51]
Experience	tourist providers' datasets (DMOs)		MinCIT-Open Data [52], DataEco [53]
Provider	government providers' datasets		MinCIT [54]
City	social networks	GeoNames	[48]
Attraction	social networks, IoT (POI wireless transmitters)	GeoNames	[48], beacons
Hotel	social networks: eWOMs, OTAs		[48]
Review	social networks: eWOMs, OTAs	time	[48]

Table 3. Glossary of a TTS (sample concepts).

Term	Synonym	Acronym	Description	Type
Attraction	Point-of-Interest	Pol	A place of interest where tourist visit for its value or significance.	Class
Tourist	Visitor		A person who travels away from their normal residential region for a temporary period of at least one night, to the extent that their behavior involves a search for leisure experiences from interactions with features or characteristics of places he/she chooses to visit.	Class
Tourist experience		TE	A set of activities in which individuals engage on their personal terms, such as pleasant and memorable places, allowing each tourist to build his or her own travel experiences so that these satisfy a wide range of personal needs.	Class
Destination	City		A geographical area consisting of all the services and infrastructure necessary for the stay of a specific tourist or tourism segment.	Class
Provider	Supplier		All businesses offering tourism services and experiences to consumers when the latter are traveling and performing tourism activities.	Class
Review	Opinion		A subjective opinion of a tourist's experience.	Subclass

Another type of relationship between classes describes their behavior. For instance, the class “Hotel” has a relationship “hasService” with “Service.” For example, from the class “Provider,” we obtained several subclasses, according to the category of the service offered, so the class “Hotel” has an include relation of “SubClass Of” from the class “Accommodation”, and this, in turn, has a relation “SubClass Of” with “Provider.” This last relationship is an illustration of ontology refinement using data-mining techniques in Big Data environments.

3.3.3. Formalization and Implementation

We used Protégé as the editor and framework for the construction of OntoTouTra. Through formalization, we produced meaningful models at the level of knowledge. We gave each class or subclass term a semantic relationship between them (see Table 4). In this phase, we solved the semantic problems detected, for instance, the need to specialize the tourist experiences in subclasses and determine subclasses for the tourist reviews according to the provider or the geographic location of the review. The formal language used was OWL/RDF.

Table 4. OntoTouTra relationships (owl:topObjectProperty).

No	Relationship	No.	Relationship
1	belongs	9	hasService
2	enjoys	10	hasServiceCategory
3	hasAccommodationType	11	hasStateParent
4	hasCityParent	12	located
5	hasCountryParent	13	offered
6	hasHotel	14	operates
7	hasHotelScore	15	uses
8	hasScoreCategory	16	visits

3.3.4. Evaluation

At this stage, we verified the level of consistency and acceptance of the ontology knowledge. We did this process from three approaches. The first consisted of verifying whether the defined objectives met the purpose of the ontology. For this, we followed the FOCA methodology. The second was the validation of the conceptual model to determine the effectiveness of the ontology. To do this, we used the CQ approach by calculating ten KPIs from a TTS system. The last approach corresponds to the test of the ontology through a use case. We generated the ontology individuals from web scraping of an OTA for the Colombian tourist case. We created ten test case scenarios with this case study and executed SPARQL for each KPI from the previous approach. The results of these SPARQL queries were contrasted with the expected results obtained from the sources of authorities in tourism. Section 5 details each of these three OntoTouTra evaluation approaches. Besides, we made a document (Supplementary Material) with the implementations and results of these test cases.

3.3.5. Documentation

The documentation is essential to recognize the current state and maintain the ontology’s consistency. For this process, we used two tools for the automatic production of the documentation: Protégé and Ontology-based APIs (OBAs) [55].

Regarding the logic model, the OntoTouTra architecture is multilayered based on functionality (see Figure 3, Section 3.4). As mentioned above, this architecture operates in Big Data environments, wherein the lower layers use data-mining techniques to process data from ubiquitous data sources. In the upper layer, the ontology offers different data recovery possibilities, such as the traditional SPARQL queries from an endpoint and REST API requests, the implementation of which can be seen in the screenshots of the Supplementary

Material. Taking advantage of these ontology query possibilities, we handled scripts in programming languages, especially Python, to perform complex queries with Big Data analytics techniques, using the PySpark and PyMongo libraries.

3.4. Model for the Development of OntoTouTra

Our model for developing the ontology of tourist traceability has the following components (see Figure 4).

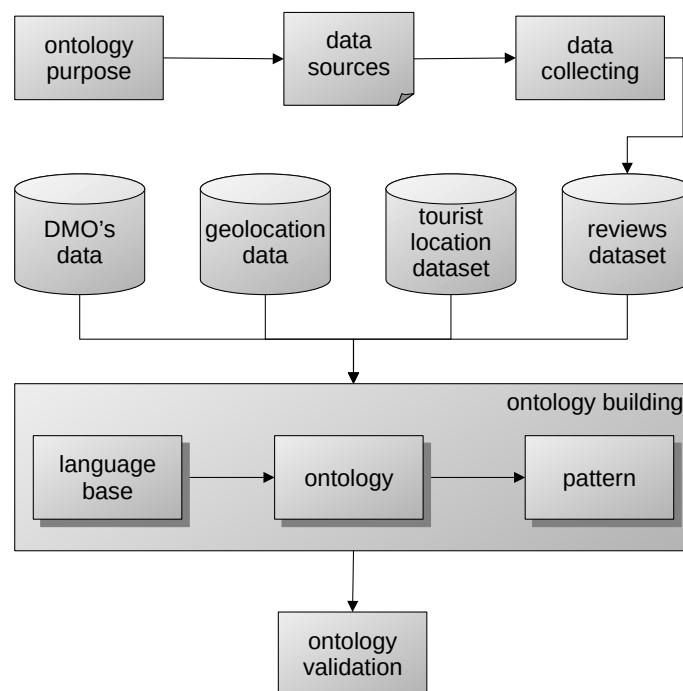


Figure 4. OntoTouTra development model.

Next, we define and explain the procedure for each of the stages of the model that we developed to create and validate the OntoTouTra ontology through lists, diagrams, tables, and statistical graphics. We also provide the necessary recommendations to satisfy the requirements of each stage.

3.4.1. Definition of the Ontology's Purpose

The model begins with the scope of the ontology of tourist traceability, the justification, the motivation, and the goals. The purpose may arise from the need for decision-making by the DMO to improve the destination and its POIs. This component is mandatory.

3.4.2. Data Sources

The sources from which the data are collected can be governmental, public, or private sources such as the regulations for the provision of tourist services, information systems, social networks, other ubiquitous sources, reports from the UNWTO, other tourism authorities, tourism reports from local and national governments, hotel occupancy data, restaurant management, and the entities that revolve around tourists (see Table 2). Given its traceability feature, geospatial data sources are significant

3.4.3. Data Collecting

We can collect data from the identified sources, which can be manual, semi-automatic, or automatic. The data can be on paper, files, datasets, ontologies, information systems, social networks, sensors, mobile devices, and the web, among others. We can use custom applications to obtain automatic or semi-automatic data, whether in batch or real-time processing (see Figure 5). Some ad hoc developments may be required, mainly to obtain

specific terms about tourism subjects that will be part of the corpus and the lexicon of the ontology, for example, to collect social networks data, API, or web scraping, then applying data-mining techniques.

BookingWebScraping
<ul style="list-style-type: none"> - locationList : list - hotelList : list - language : string - location : string - directory : string - locationHeader : string - hotelHeader : string - source : string - inputFilename : string - country : string - countryName : string - web : string - cityKeyword : dict - hotelKeyword : dict - propertyTags : dict
<ul style="list-style-type: none"> + getLocation(region : string, language : string) : list + getPropertiesReview() : list + addLocation(hotelID : int, hotelName : string, hotelURL ... + getPropertiesList(locationID : int, location : string) : list + getLocationID(location : string) : int + hotel2CSV(reviews : list) + location2CSV(location : string, reviews : list) + getLocationReviews(location : string) : list + getSource() : string + getLocationsByCountry() : list + getPropertySource() + getPropertiesByCountry() : list + getCountryReviews()

Figure 5. Web scraping class.

3.4.4. Tourist Location Dataset

Tourist traceability requires the tracking of their geographical location. The calculations of the geographic positions, the permanence in the POI, and the destination can be performed by utilizing coordinates or even by semantic analysis, which determines a specific location. Therefore, the classes of the ontology must have subclasses or attributes that facilitate the determination of geographic coordinates (see Figures 6 and 7). Ontologies and external geographic datasets can form this component. The ontology must interpret the terms of locations, mainly as nouns or names.

```

PREFIX ott: <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX gn: <http://www.GeoNames.org/ontology#>
PREFIX wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace/>

SELECT ?city ?cityName ?stateName ?countryName
       ?geo ?geoName ?alternateName ?countryCode
       ?cityLat ?cityLon
WHERE {
  ?city    ott:cityID           ?cityID           ;
          ott:cityName        ?cityName          ;
          ott:stateID         ?stateID           ;
          ott:hasStateParent  ?state             .
  ?state   ott:stateName       ?stateName        ;
          ott:hasCountryParent ?country          .
  ?country ott:countryName     ?countryName     .
  ?geo     gn:name             ?geoName          ;
          gn:alternateName     ?alternateName    ;
          gn:countryCode       ?countryCode     ;
          gn:population        ?cityPopulation  ;
          wgs84_pos:lat        ?cityLat         ;
          wgs84_pos:long       ?cityLon         .
  FILTER(?cityName = "Cartagena de Indias") .
  FILTER(CONTAINS(?alternateName, ?cityName)) .
  FILTER(LANG(?alternateName) = "es") .
}
LIMIT 1

```

Figure 6. Listing of Data link to GeoNames for obtaining city coordinates.

```

{
  "head": {
    "vars": [
      "city", "cityName", "stateName", "countryName", "geo",
      "geoName", "alternateName", "countryCode", "cityLat", "cityLon"
    ]
  },
  "results": {
    "bindings": [
      {
        "alternateName": {
          "type": "literal",
          "value": "Municipio de Cartagena de Indias",
          "xml:lang": "es"
        },
        "city": {
          "type": "uri",
          "value": "http://tourdata.org/ontotoutra/ontotoutra.owl#CartagenaDeIndias"
        },
        "cityLat": { "type": "literal", "value": "10.50743" },
        "cityLon": { "type": "literal", "value": "-75.4543" },
        "cityName": { "type": "literal", "value": "Cartagena de Indias" },
        "countryCode": { "type": "literal", "value": "CO" },
        "countryName": { "type": "literal", "value": "Colombia" },
        "geo": { "type": "uri", "value": "https://sws.GeoNames.org/3687247/" },
        "geoName": { "type": "literal", "value": "Municipio de Cartagena de Indias" },
        "stateName": { "type": "literal", "value": "Bolívar" }
      }
    ]
  }
}

```

Figure 7. Results of data link to GeoNames for obtaining city coordinates.

As can be seen in the results of the query listing, there are no terms for latitude and longitude within the terms of the OntoTouTra ontology for cities. Using a data link to GeoNames, these terms are obtained. To avoid ambiguities with the names of the locations, we attempted to obtain from the ubiquitous data source the most significant amount of data that characterized the location of that geographic entity, for instance the type of unit: country, state or region, city, municipality, and neighborhood, among others; geographic coordinates (latitude and longitude) and direction. Furthermore, we established the relation of ontological classes, for example, a city has the relationship “hasStateParent,” a hotel has the relationship “hasCityParent,” and so on.

In tourism traceability, queries on the geographical issue are needed, which OntoTouTra alone would not solve. GeoNames is a specialized ontology and is ideal for complementing geographic data that OntoTouTra lacks., for instance, to perform population-related calculations, such as the rate of tourism companies for every number of inhabitants. OntoTouTra makes linked data with GeoNames and retrieves the data of the number of inhabitants of a specific geographic area.

3.4.5. Tourist Reviews Dataset

The tourist reviews provide items for the ontology; they are terms frequently used in tourist slang and the valuable channel of communication and feedback for the tourist ecosystem. Reviews can be obtained manually, such as surveys and suggestion boxes, or automatically extracted from tourism social networks depending on the data source.

Forming a dataset of tourist reviews has many advantages and serves as a corpus of the ontology. For example, through NLP, we can obtain the polarization of the reviews. We can also establish the traceability relationship, that is spatiotemporal. In Figure 8, we see the distribution of the scores the tourists gave to the localities (cities) that they visited, through the process and visualization of the dataset of tourist reviews of Colombia, in English. In addition, through NLP, we can tokenize the tourist reviews, and in this way, the ontology terms are achieved through a filter. NLP also allows classifying the terms. This component can enrich it with unsupervised-machine-learning techniques to cluster the terms. Some terms of the previous component may be wrongly spelled, poorly categorized, or not relevant to the ontology developed. We used a simple filtering method to determine the frequency of valid terms accepted by the ontology. More filters can be applied to search the quality of the corpus of the ontology.

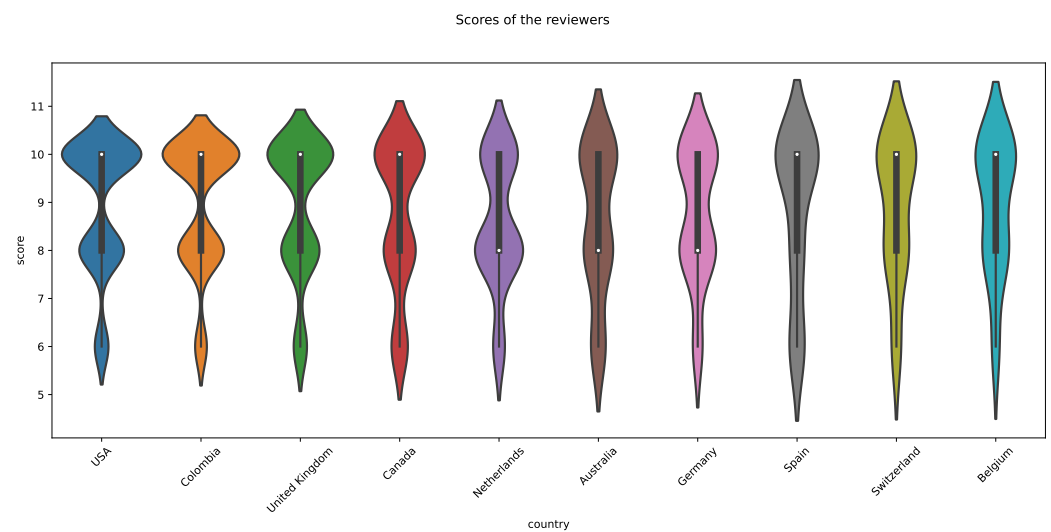


Figure 8. Distribution of the scores of the top 10 nationalities of reviewers of Colombia’s tourist reviews dataset obtained from OntoTouTra (language: English).

3.4.6. Ontology Input Data Files

We entered individuals into the ontology manually or automatically. The current version of OntoTouTra was designed in Protégé [56], using the Cellfie plugin [57]. We uploaded the individuals’ spreadsheets. Cellfie creates the axioms of the ontology, using transformation rules, as seen in Figure 9. Regarding the two remaining data sources, and as we mentioned earlier, the ontology design recommends using ubiquitous data with Big Data analytics techniques. Class instances such as tourism experiences and provider data are often in these formats and can be loaded into the ontology. Whenever possible, we recommend reusing knowledge through open link data for geolocation data, which is very sensitive for a traceability system.

```
{
  "Collections": [{
    "sheetName": "city",
    "startColumn": "A",
    "endColumn": "A",
    "startRow": "2",
    "endRow": "+",
    "comment": "",
    "rule":
      "Individual: @B*\n
      Types: City\n
      Facts: \n
      cityID @A* (xsd:integer),\n
      cityName @B*,\n
      stateID @C* (xsd:integer)",
    "active": true
  }]
}
```

Figure 9. An example of transformation rules from the Cities spreadsheet.

3.4.7. Ontology Building

Base language -> ontology -> pattern. This is closely related to the first three of the five proposed phases (Specification, Conceptualization, Formalization, Implementation, and Maintenance) of the METHONTOLOGY methodology [46,58]. The OntoTouTra ontology architecture is multilayered (see Figure 3) based on functionality, from storage (low-tier) to interaction (top-tier).

- Layer 1 corresponds to the input data, mainly from ubiquitous computing sources, such as social networks, sensors located at the destination, and users’ mobile devices. This process was carried out through a data analysis pipeline, where we applied

qualitative and quantitative techniques when examining the data to provide valuable insight. Data analytics provides the means to examine the EDA and CDA findings. Using EDA, we explored the data to find patterns and relationships among different ontology elements. Furthermore, through CDA, we obtained conclusions to specific questions of the tourism domain, based mainly on the simple observation of the data.

- Layer 2 is the logical layer, achieved by reasoning from OWL/RDF storage. The reason is limited according to the domain and range restrictions defined in the ontology. Using this layer, we can explain the content, apply queries, and verify the integrity of the ontology.
- Layer 3 corresponds to the presentation; OntoTouTra allows data visualization with different SPARQL endpoints, APIs, and graph visualization tools.

3.4.8. Ontology Validation

This ensures that the ontology fulfills its purpose (first component). Steiner and Albert [59] suggested the validation of the content, application, and structure. The ontology must work appropriately according to its approach with the criteria of consistency, completeness, and conciseness. The validation of the functional ontology was performed on a set of domain CQ tests. The tests were implemented as queries of the individuals of the ontology. These tests were confirmed with the reasoning system. OntoTouTra uses the Protégé reasoning system, as is the case with Hermit Version 1.4.3.456 [60].

4. Development and Usage of OntoTouTra in Big Data Environments

Big Data is part of a strategic initiative to design and execute business technology solutions backed by the analysis and management of large volumes of data through technology [16]. Big Data is an ideal solution for analyzing, processing, and storing data from tourist traceability systems. We needed to combine multiple unrelated datasets, analyze the data provenance, process large amounts of unstructured data, and look for hidden data patterns in a time-sensitive way. The analysis allowed understanding the data, examining it employing scientific techniques and automated tools to discover hidden behaviors and patterns. From massive amounts of data, without processing or structuring, the relevant information was obtained. A methodology is needed to handle the different requirements to execute Big Data analytics.

Ubiquitous data sources, such as social networks and the IoT, require massive parallelism to obtain the vast volumes of data, the data distribution, high-speed networks, and data mining and analytics. A tourist traceability system depends on the processing of these data; we were interested in knowing the activity of the tourist within the destination and its relationship with the tourist actors. The reviews are an excellent example of this interaction since they provided us with that fundamental space–time causality for traceability. An alternative to analytics is graph analytics, which uses an abstraction called a graph model. This model connects large volumes of data from different sources and in various structures. Graph analytics gather structured and unstructured data by coupling them into entity relationships. We can infer, identify patterns of interest, and deduce through an iterative approach to discover knowledge through this analysis. The same as the ontology, the graph model is straightforward since it is based on entities (nodes) and edges (relationships) [61].

4.1. Big Data Analytics Lifecycle for Building the TTS Ontology

This methodology allows planning and organizing the tasks, activities, and resources for data management. As a methodology, this research adopted the lifecycle of data analytics [16], divided into nine states (see Figure 10).

In 2016, Erl et al. proposed a lifecycle model for Big Data analytics [16]. It is a step-by-step methodology necessary to organize the activities involved in the acquisition, processing, analysis, and reuse of data. This methodology is applicable in any context. For this reason, we adapted these methodological phases for the construction and use of OntoTouTra in Big Data environments. Next, in each of the stages, we explain this

adaptation, and employing some lists and charts, we demonstrate the implementation that we carried out in the ontology.

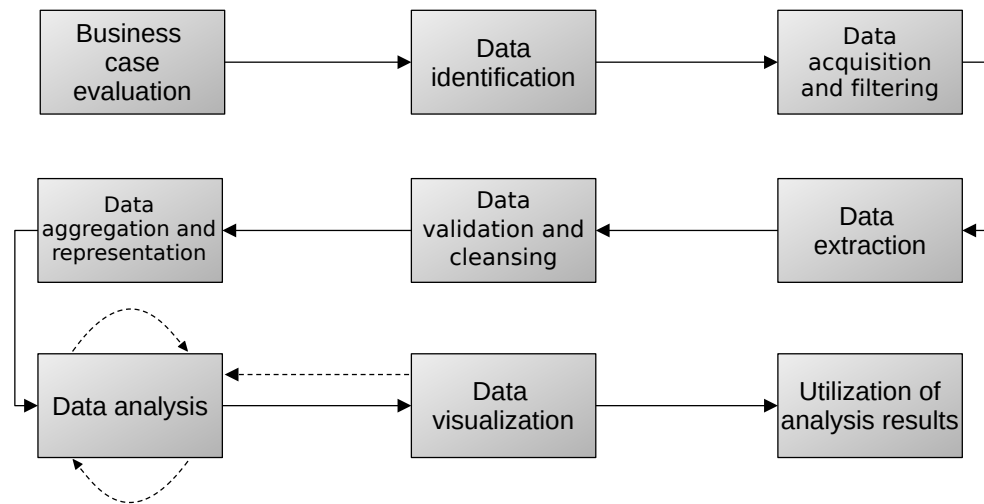


Figure 10. Big Data lifecycle [16].

4.1.1. Business Case Evaluation

This is the stage related to the first and last component of the OntoTouTra model. It is necessary to have clarity about the justification, the motivation, and the objectives of the tourist traceability analysis. The motives to carry out this analysis can be various, among which we can mention: the marketing domain and the destination promotion, the actors involved in tourism management, the definition and application of policies and strategies, destination management, decision-making, and financial management [62] (see Figure 1). It is necessary to seek advice from expert Big Data and tourism management consultants because not all solutions meet the conditions and features of Big Data (the 5Vs: volume, variety, velocity, value, and veracity).

4.1.2. Data Identification

In this stage, we determined the datasets and provenance. Location and tracking data for tourists within the destination are indispensable to satisfy the requirements of the previous step. Around the data, we required establishing the acquisition cost, confidentiality, and personal data treatment policies. Table 5 shows the leading OTAs that are potential data sources with information on the tourist domain. Booking.com registers the most significant number of accommodation listings for tourist site information and a tourist review platform. For many years, it has remained in the top 10 of the OTAs with the most excellent offer. In particular, in our case study, we chose this OTA.

Table 5. OTAs (source: Cloudbeds, 2020).

OTA	Founded	Listings	Audience	Countries	Languages
Booking.com	1996	28 M	50 M	200	43
Skyscanner	2001	2 M	60 M	49	30
Expedia	1996	590 K	50 M	75	35
TripAdvisor	2000	7.3 M	490 M	48	28
Agoda	1998	2 M	2.3 M	65	38
Airbnb	2008	7 M	750 M	220	89
HostelWorld	1999	36 K	13 M	178	20
Hotelbeds	2001	180 K	60 K	185	20

4.1.3. Data Acquisition and Filtering

This involves gathering data by different means: files, digitalization, web scraping, integration with API, cloud services, transactional data, sensor data, information systems databases, and dataset providers, among others. Filtering is necessary to eliminate noise from the data. It is desirable to use data-mining techniques. In Figure 11, we see the Python class that invokes the Selenium Web Scraping driver. The routes and parameters of the OTA were previously defined. Web scraping is hierarchically performed by region; for instance, we can start with a specific country and then go through its states or subregions. We also designed the class methods to obtain the information in a structured way from the hotels: general info, address, services, ratings, and reviews.

```
# Selenium web scraping driver
d = webdriver.Chrome(chromeDriverDir)

# Iterate states of a country
for state in states:
    hotelsDF = getHotelData(d, state, ' ')

    urlNextPage, hotelsPerPage, totalPages, hotelsNum = getNextPageURL(d)
    startInstance = len(hotelsDF) if hotelsPerPage == 0 else hotelsPerPage
    while (startInstance + 1 <= hotelsNum):
        newURL = urlNextPage + tags['offset']['tag'] + str(startInstance)
        hotelsDF = hotelsDF.append(
            getHotelData(d, state, newURL),
            ignore_index = True
        )
        startInstance += hotelsPerPage

    hotelsDF = getHotelAddress(d, hotelsDF)
    hotelsDF, hotel_services = getHotelServices(d, hotelsDF)
    hotelsDF, hotel_reviews, hotel_scores = getHotelComments(d, hotelsDF)
    hotelsDF.drop_duplicates(subset = 'hotelID', inplace = True)
    hotelsDF.reset_index(drop = True)
    servicesDF = pd.DataFrame(hotel_services)
    servicesDF['hotelID'] = hotelsDF['hotelID']
    scoresDF = pd.DataFrame(hotel_scores)
    scoresDF['hotelID'] = hotelsDF['hotelID']

# Get the keys of the dictionary list
list(dict.fromkeys([k for i in hotel_reviews[0] for k, v in i.items()]))
# Get the hotel reviews Dataframe
hreviews = []
for hotel in hotel_reviews:
    for comment in hotel:
        hreviews.append(comment)
reviewsDF = pd.DataFrame(hreviews)
```

Figure 11. Python code snippet about OTA web scraping.

4.1.4. Data Extraction

This extracts disparate data and transforms then into an understandable format for the Big Data solution. In the case of scanned documents, at this stage, it is determined if the Big Data solution can read them in their original format or if we need to execute OCR applications. In the final part of Figure 11, we see how the data are extracted and stored in memory in the datasets that were formed for each of the data structures of the hotels, tourist destinations, and tourist reviews.

4.1.5. Data Validation and Cleansing

Validation rules and removal of invalid data are applied to determine the accuracy and quality, for instance, the validation of destination geographic coordinates and the tourist activity timestamps. This stage is very demanding in a web scraping operation because the way the data are displayed on the OTA web pages can vary. Often, the information is missing, erroneous (due to user typing), or may be intermittent due to the conditions of Internet access to the site.

4.1.6. Data Aggregation and Representation

We required a unified data view by identifying the key fields to join sparse datasets because they come from different sources. This is a complex process because the syntax and semantics of the data model are determined. We designed this model with reuse principles for future requirements.

4.1.7. Data Analysis

We set the ontology axioms and terms. Different types of analytics were applied to discover data patterns through operations such as queries, aggregations, or filters. The analysis can be confirmatory or exploratory, depending on the deductive or inductive approach. When we supply the ontology individuals (instances), exploratory analysis is the most suitable because it is closely related to data mining.

4.1.8. Data Visualization

The results' interpretation leads to the formulation of the ontology, determining its structure (classes, relationships, functions, axioms or restrictions, instances, and properties or attributes), hierarchy, clarity, extensibility, and coherence. An example of the visualization of our ontology can be seen in Figures 12 and 13. A SPARQL geolocated query was executed on the ontology. Later, we stored it in a dataset, and using the plotly.express library, we visualized the results of the query on a map;

```
import pandas as pd
import rdflib

g = rdflib.Graph()
onto_filename = os.path.join(path, 'ontotoutra.owl')
format_ = rdflib.util.guess_format(onto_filename)
g.parse(onto_filename, format=format_)

qres = g.query(' '
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ott: <http://tourdata.org/ontotoutra/ontotoutra.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?hotelID ?cityName ?stateName ?score ?lat ?lon
WHERE {
    ?hotel ott:hotelID ?hotelID ;
           ott:hotelLat ?lat ;
           ott:hotelLon ?lon ;
           ott:hotelReviewScore ?score ;
           ott:hasCityParent ?city .
    ?city ott:cityID ?cityID ;
          ott:cityName ?cityName ;
          ott:hasStateParent ?state .
    ?state ott:stateName ?stateName .
}' ')
city_data = []
for hotel, city, state, score, lat, lon in qres:
    if len(city.value) > 0:
        city_data.append([hotel.value, city.value, state.value,
                          score.value, lat.value, lon.value])
city_df = pd.DataFrame(city_data,
                       columns = ['Hotel', 'City', 'State', 'Score', 'lat', 'lon'])

dir_path = os.path.dirname(os.path.realpath(__file__))
token_filename = os.path.join(dir_path, '.mapbox_token')
token = open(token_filename).read() # token from Mapbox

import plotly.express as px
fig = px.scatter_mapbox(
    city_df, lat='lat',lon='lon', hover_name='City',
    hover_data=['State', 'Score'], color_discrete_sequence=['fuchsia'],
    zoom=5, height=1000
)
fig.update_layout(mapbox_accessToken=token)
fig.update_layout(margin={'r':0, 't':0, 'l':0, 'b':0})
fig.show()
```

Figure 12. Example of ontology visualization: Main tourist destinations in Colombia.

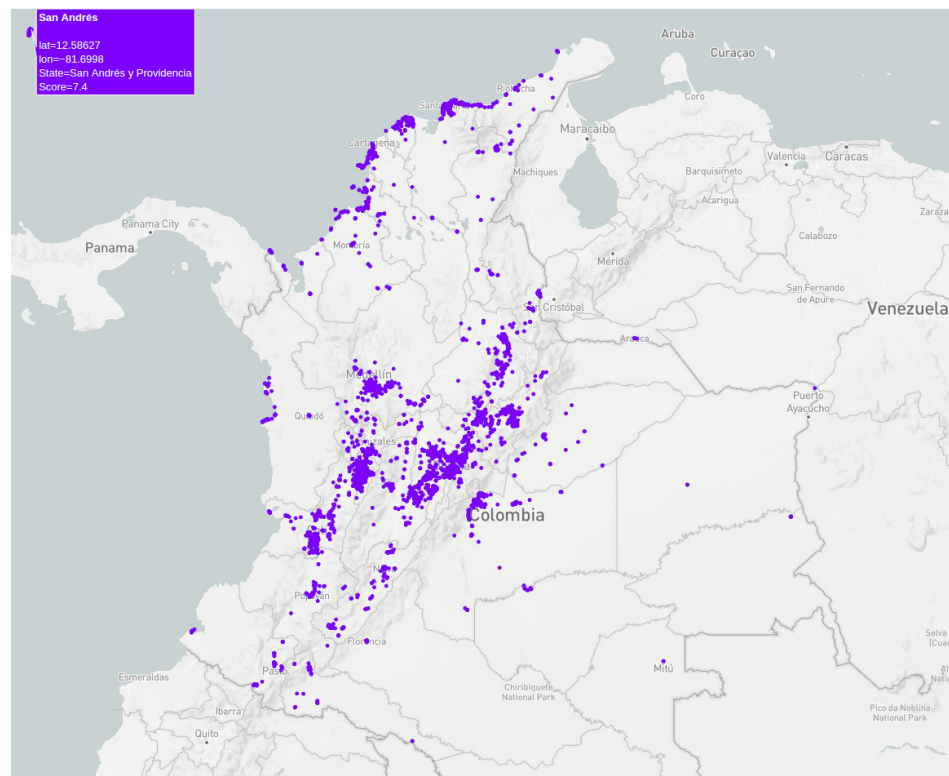


Figure 13. Example of the visualization of tourist destinations in Colombia from OntoTouTra.

4.1.9. Utilization of Analysis Results

We built the OntoTouTra ontology, whose primary purpose is the knowledge base for a tourist traceability system. The results of the analysis can support decision-making for the tourism ecosystem. For example, we can apply machine learning and NLP techniques to determine the KPI of tourist satisfaction at the destination based on their reviews (see Figures 14 and 15). In Figure 16, the polarity corresponds to the x-axis and the subjectivity to the y-axis. Polarity determines whether the review is positive or negative, while the size of the chart markers determines the subjectivity. We found more positive reviews located on the right-hand side.

```

from nltk.sentiment.vader import SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()

reviews['neg'] = reviews['review'].apply(lambda x:sia.polarity_scores(x)['neg'])
reviews['pos'] = reviews['review'].apply(lambda x:sia.polarity_scores(x)['pos'])
reviews['neu'] = reviews['review'].apply(lambda x:sia.polarity_scores(x)['neu'])
reviews['compound'] = reviews['review'].apply(
    lambda x:sia.polarity_scores(x)['compound'])

from plotly.offline import download_plotlyjs, init_notebook_mode, plot
import plotly.express as px

reviews['YearMonth'] = pd.to_datetime(reviews.date).apply(
    lambda x: '{year}-{month:02d}'.format(year=x.year, month=x.month))

df = reviews.groupby(reviews.YearMonth)[['pos', 'neg', 'neu']].mean()
df.reset_index(inplace=True)
df = df.rename(columns = {'YearMonth':'date'})
fig = px.line(df, x=df.date, y=[df.pos], hover_data={'date': "|%B %d, %Y"},
              title='Satisfaction KPI')
fig.update_xaxes(dtick="M1", tickformat="%b\n%Y",
                 ticklabelmode="period")
plot(fig)
  
```

Figure 14. Application of sentiment analysis techniques to determine the Satisfaction KPI in Colombia.

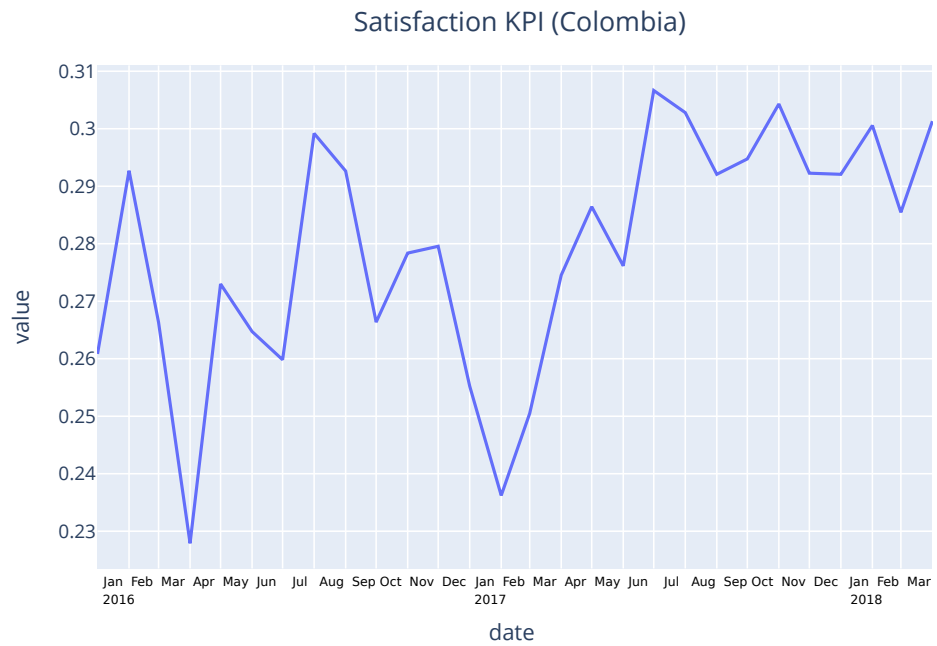


Figure 15. Example of satisfaction KPI (Colombia): positive reviews of the destinations. Obtained from OntoTouTra.

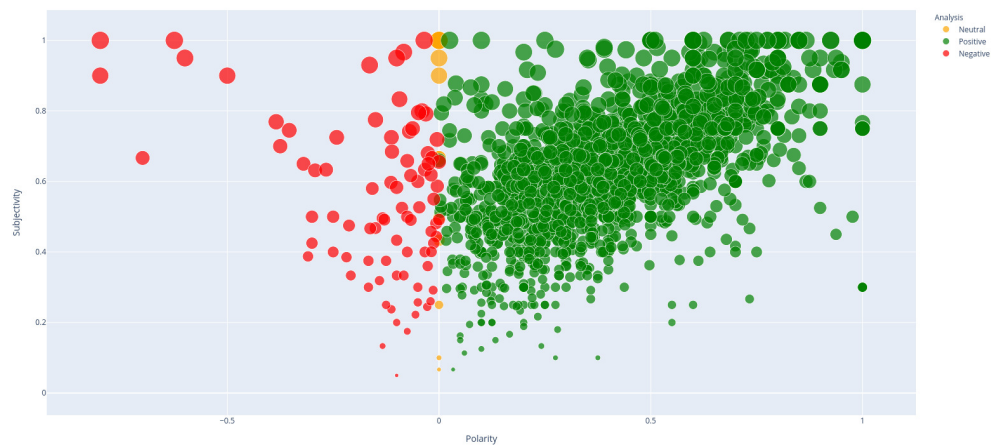


Figure 16. Example of the polarity and subjectivity of the reviews about the Colombian destinations. Obtained from OntoTouTra.

4.2. Using Big Data

4.2.1. Components of the Analytics Toolkit

In this study, we utilized some key Big-Data-mining technologies to define the classes and terms of the ontology and build some queries. Table 6 shows the analytics toolkit used in this research.

The architecture diagram of the data pipeline can be seen in Figure 17. In the case of this diagram, we started by obtaining the data from a ubiquitous data source from an OTA (Booking.com), using web-scraping techniques with the Selenium library in Python. Then, we created the data flow of the respective data unit according to the scraping; we worked with the destination and its geographical coordinates, the data of the suppliers, especially the hotels, the tourist services, their ratings, and the tourist reviews with their temporality data. These streams were written as documents to a MongoDB collection. Subsequently, we built a Spark Streaming Dataframe that reads the MongoDB collection and periodically updates or adds new data. We made structured queries from the Spark Streaming Dataframe to store their results as axioms in the ontology. These axioms are of

two types: The first type corresponds to detecting new patterns of data units that boost the ontology with new classes or terms (for example, a new attribute for the class “Provider” or a new class representing a tourist actor within the ontology). The detection of new data units for the ontology was carried out with NLP applied to the tourist reviews. The second type of axiom corresponds to the generation of new individuals in the ontology, such as, for instance, the creation of a new tourist experience, new groups of reviews, new hotel instances, or new tourist destinations or POIs.

Table 6. Components of the analytics toolkit.

Software	Use	Function
Spark/PySpark	data mining	PySpark Dataframe for Big Data entities: reviews, hotel services, and scores.
MongoDB	data mining	Temporary storage for NoSQL collections, mainly tourist reviews.
Python	data mining/queries	Scripting for all functions: scraping, ontology API, loading of individuals, queries, and visualization.
RDFLib	queries	SPARQL API interface.
Selenium	data mining	OTA web scraping.
NLTK	data mining	Definition of ontology classes and terms. Analysis of tourist reviews for queries.

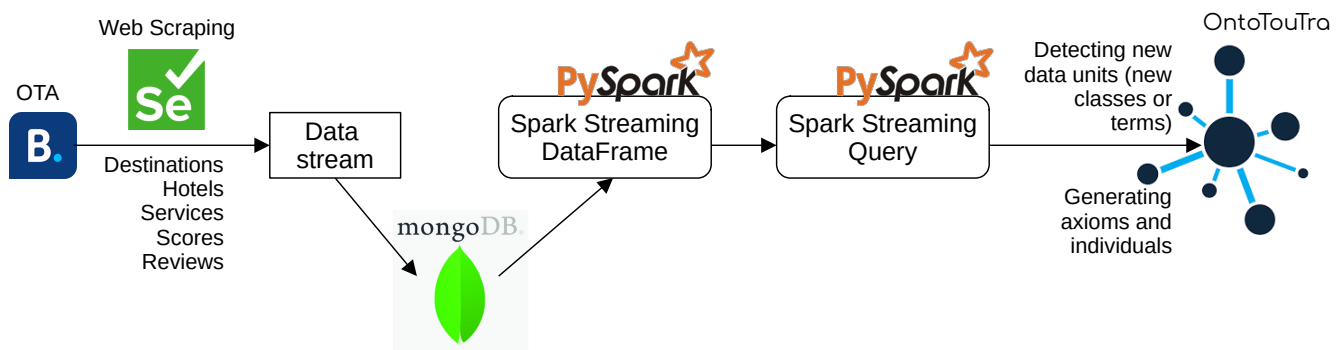


Figure 17. Architecture diagram for the data pipeline.

To carry out complex queries that require the calculation of an enormous amount of data, we also used Big Data, for example, in the stages of data aggregation and representation, data analysis, data visualization, and the use of the analysis results, explained in Section 4 and Figures 10, 12 and 13; here, we executed the scripts on the OntoTouTra ontology and visualized the tourist destinations of Colombia on a georeferenced map.

4.2.2. Variety of Data

When applying web scraping, we collected data of various structures and, in some cases, without structure. In Figure 18, we see the data flow of a tourist review in HTML code. We used MongoDB because it is a NoSQL database that stores unstructured data in the form of documents. In this way, using a Python script, we analyzed the data flow obtained from web scraping and converted it into JSON format for later loading into MongoDB, that is, we went from unstructured data to semistructured data. Subsequently, with the Streaming Dataframe and the Streaming Query, we generated the axioms of the individuals of the ontology in data in a structured way.

- The identification of relevant terms from a large and messy data source. Web-scraping techniques allowed obtaining, cleaning, and filtering the data from the tourist social networks sites. Due to the volume, variety, and velocity features, Big Data pipelines were designed and implemented for data processing;
- Significance and value of the domain. NLP techniques were applied to filter the terms to build the knowledge base of the ontology;
- Ontology construction: Big Data provided facilities for the data preprocessing so that later, an ontological building tool facilitated the creation of the thesaurus, the classifications, the taxonomy, the concept sets, the link between concepts, documentation, grouping in collections, mapping employing concept schemes, inference, and mapping link;
- The reasoning. The bidirectional relationship of Big Data semantics was fundamental in the application of the OntoTouTra ontology. The semantic basis was the ontology. For instance, we set axioms that determined the polarity of the tourist reviews.

The tourist reviews from the OTA gathered through web scraping became the ideal input to apply Big Data analytics because these fulfilled its features. The tourist reviews offered various aspects concerning the domain of the TTS, such as the location, time, services, ratings, and of course, the opinion. We extracted these aspects with opinion-mining techniques, and we had challenges such as the classification of multi-aspect opinions. We identified the vocabulary from these reviews using the NLP as a first step to face this challenge. We used machine learning classification methods and, in some cases, deep learning.

Besides the construction of the ontology, we also used Big Data analytics for its use, for instance, in the data visualizations such as Figure 13 and the predictions such as the review scores depicted in the algorithm of Figure 19. We used this algorithm for a double function: to generate the ontology vocabulary corpus and, in turn, to predict ratings. The algorithm preprocessed the data from the reviews. Specifically, in the data cleaning, we used the lemmatizing of the reviews and other NLP techniques such as tokenizing by word and by sentence, filtering stopwords, stemming, and tagging. To this end, we worked on Python libraries such as SpaCy, NLTK, Tokenizer, and Keras pad_sequences. We were also able to identify the language of the review. In the case of English, using this algorithm, we formed a vocabulary size of 16,466 terms and a maximum sentence length of 197 characters from a dataset of 57,063 instances. Each instance had a positive or negative review or both. The analysis of this vocabulary defined the classes with their attributes and their relationships. This definition was checked with the sources of the tourism authorities to enrich the definition of the ontology. The vocabulary was obtained in the sixth step, "Keras_create_vocabulary." The remaining steps of the algorithm were intended to generate the model, train it, and predict the rating of the reviews as an application of the use of the ontology. Figure 20 shows the validation results of the prediction of this algorithm in both loss and accuracy. Reasonable results can be seen, although low prediction. A model based on a bidirectional long short-term memory (LSTM) network and four fully connected layers was used. We could reduce the overfitting further by increasing the dropout layers of this deep-learning model.

4.2.4. Classification Using Big Data

Big Data analytics describes data, control technologies, analysis methods, and data mining development [65]. OntoTouTra's data sources are ubiquitous, primarily social networks. We used data mining and Big Data analytics as a decision support process by searching raw data for hidden patterns that are useful and interpretable for decision-making in the TTS domain. In this way, we extracted facts and generated hypotheses using statistical tools, artificial intelligence, and machine learning.

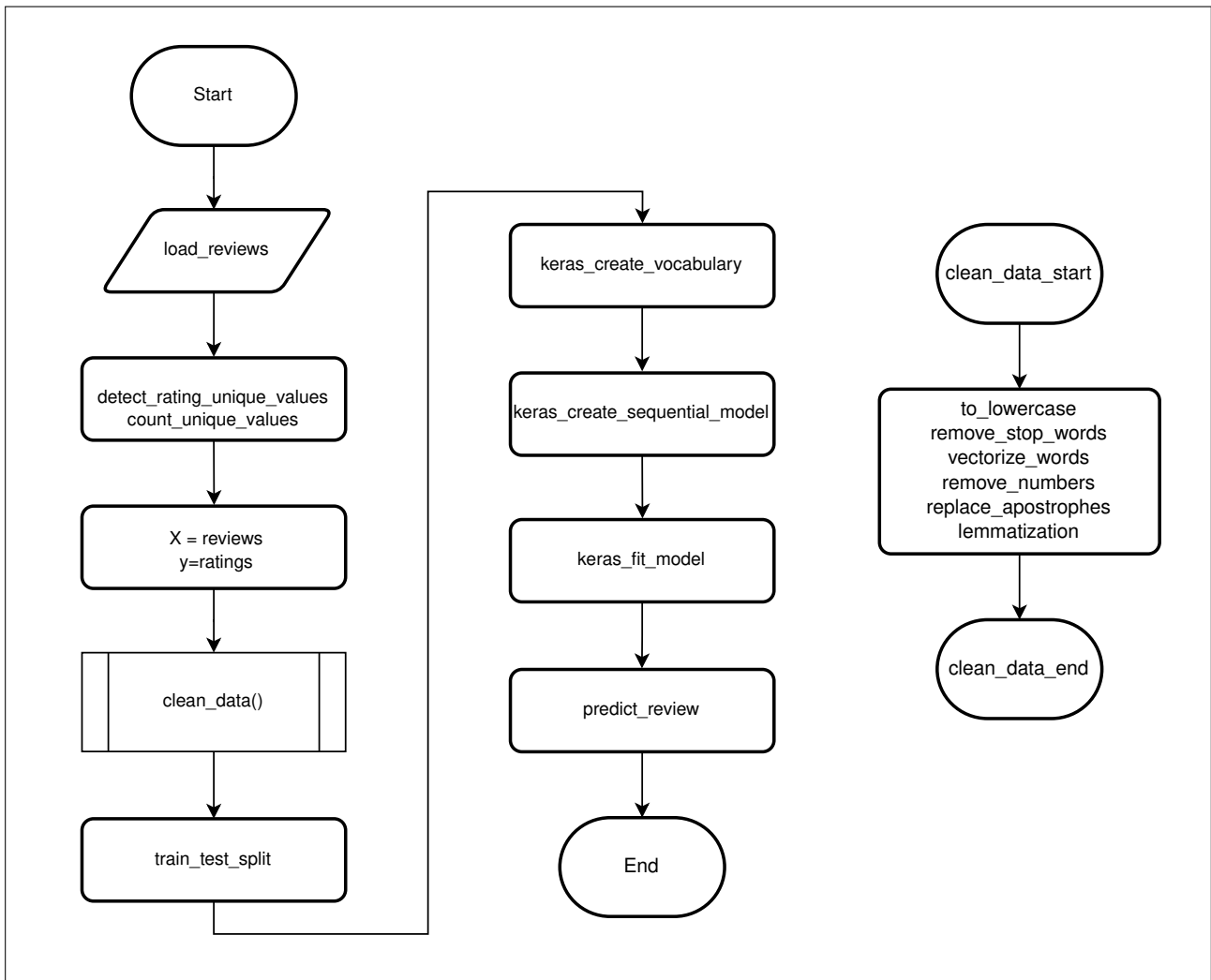


Figure 19. Rating predictor algorithm.

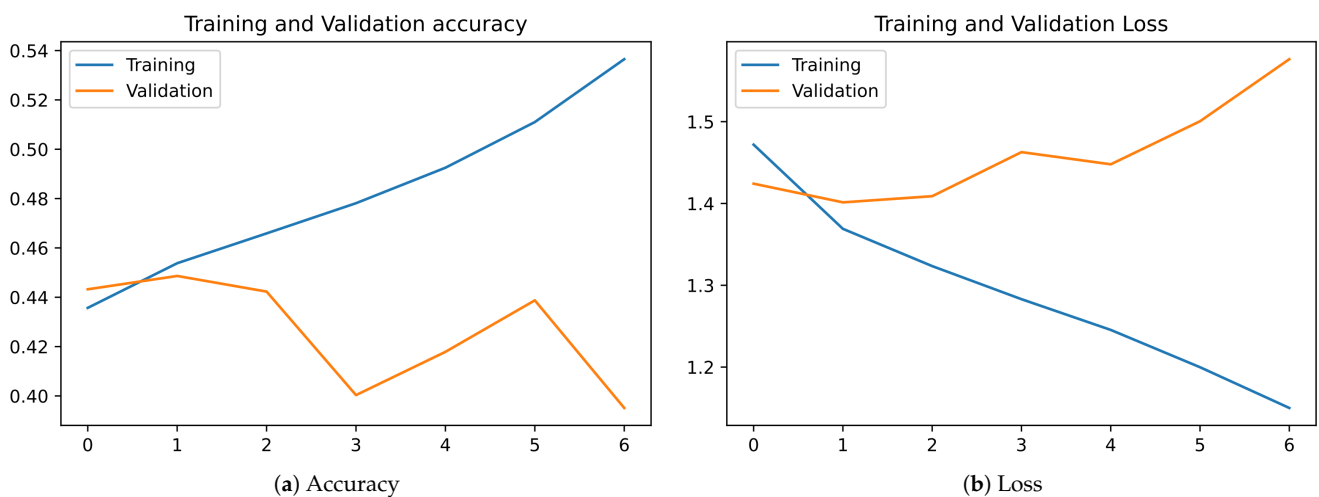


Figure 20. Performance of the rating prediction model.

We found the use of Big Data beneficial for processing structured, semistructured, and unstructured data (see Section 4.2.2) due to the web scraping applied to an OTA because data, especially tourist reviews, are characterized by the Big Data 3V requirement (volume, velocity, and variety).

The Big Data analytics applications for this study are synthesized as follows:

- Refinement of the ontology: A vocabulary was generated with NLP techniques (see Section 4.2.3) to obtain the glossary of the TTS domain to implement the stages of the specification and conceptualization of the ontology (see Section 3.3, Table 3);
- Data validation and cleaning: Using data-mining and text-mining techniques, we applied text preprocessing to the tourist reviews (see Sections 4.1 and 4.2.3 and Figure 19), such as tokenization to obtain terms by removing spaces in blank and other punctuation symbols; removal of numbers so as not to affect the review sentiment measurement; elimination of stopwords; removal of scores; stemming according to language; and applying filters to determine the effect of a denial;
- Classification of reviews: The reviews provided us with different categories of data, and based on these categories, we were able to classify them. Not all categories were present in a review. Depending on the category, we applied supervised- and unsupervised-machine-learning classification algorithms. Table 8 depicts the categories identified in the reviews and the type of classification algorithm used depending on whether the reviews had labels;
- Prediction of reviews rating: We used a bidirectional-LSTM-network-based classifier to predict ratings using the vocabulary generated from the review terms (see Section 4.2.3 and Figure 19);
- Data visualization: Using the programming and processing model, MapReduce, we generated Big Data datasets with a distributed and parallel algorithm on a cluster. We used the map procedure to filter and sort the displayed data, and we executed the summary operations with the reduce method. An example is the heat map visualization in Figure 13, where we mapped the country's regions and reduced the hotels count by region to represent them on a map with the `plotly.express` library.

Table 8. Tourist review categories.

Category	Classifier	Algorithm or Tool
Determine the polarity	Supervised	<code>nltk.sentiment.sentiment_analyzer</code>
Grouping by ratings	Not supervised	K-means
Detection of services	Supervised	Named entity recognition (NER) with SpaCy
Detection of tourist experiences	Supervised	NER with SpaCy
Detection of POIs	Supervised	NER with SpaCy
Detection of language	Supervised	<code>nltk.stem</code>

5. Evaluation

5.1. Evaluation of the Ontology

In evaluating the ontology, we verified whether the objectives defined in the “purpose of the ontology” stage were met and verified whether the ontology was built correctly. We considered the quality criteria proposed by Gruber [66]: clarity, coherence, extensibility, minimal coding bias, and minimal ontological commitment, as the evaluative metrics of the ontology. First, we checked the internal consistency of the ontology; we used the Hermit reasoning [67] tool, included in Protégé. Once this reasoner was executed, no semantic, infinite loops or partition errors were found. As a second tool, we used OOPS! [68] to detect pitfalls in the ontology, which listed a minor pitfall related to the URI containing the file extension “.owl.” As a minor suggestion, we skipped this pitfall.

Then, we used the GQM approach of the FOCA methodology [69], consisting of the thirteen questions observed in Table 9. The objective of this approach is to verify the domain and application of the ontology.

Table 9. Applying the goal–question–metric approach from the FOCA methodology on the TTS ontology domain.

Goal	Question	Metric	Note	Question Grade	Goal Grade
1. Check if the ontology complies with substitutes	Q1. Were the competency questions defined?	Completeness	13 KPIs as CQ	100	83.3
	Q2. Were the competency questions answered?	Completeness	13 KPIs answered	100	
	Q3. Did the ontology reuse other ontologies?	Adaptability	Open link data with GeoNames and Time Ontology	50	
2. Check if the ontology complies with ontological commitments	Q4. Did the ontology impose a minimal ontological commitment?	Conciseness	Ontology uses abstractions to define concepts	75	75
	Q5. Did the ontology impose a maximum ontological commitment?	Conciseness	Ontology does not use many primitive concepts	-	
	Q6. Are the ontology properties coherent with the domain?	Consistency	Checked by HermiT reasoning (Protégé plugin)	75	
3. Check if the ontology complies with intelligent reasoning	Q7. Are there contradictory axioms?	Consistency	Checked by HermiT reasoning (Protégé plugin)	100	100
	Q8. Are there redundant axioms?	Conciseness	Checked by HermiT reasoning (Protégé plugin)	100	
4. Check if the ontology complies with efficient computation	Q9. Did the reasoner bring modeling errors?	Computational efficiency	1 minor error; Checked by OOPS!	75	75
	Q10. Did the reasoner perform quickly?	Computational efficiency	Depending on Protégé capacity (we ran without the reviews' individuals: 17.197 ms)	75	
5. Check if the ontology complies with human expression	Q11. Is the documentation consistent with modeling?	Clarity	Documentation generated by Protégé	100	100
	Q12. Were the concepts well written?	Clarity	We used the ontology annotations (rdfs:comment)	100	
	Q13. Are there annotations in the ontology that show the definitions of the concepts?	Clarity	We used the ontology annotations (rdfs:comment)	100	

The FOCA methodology is ideal for evaluating ontologies based on the GQM approach for an empirical evaluation, knowledge representation roles, and metrics based on the evaluation criteria. After iteration or in total, the GQM approach is executed, and finally, the quality of the ontology is calculated. First, the ontology validation must consider the type of ontology, whether a domain, task, or application. In the case of OntoTouTra, we think it is an application ontology because the concepts are described depending on a particular domain and task, in our case the TTS, which are specializations of related ontologies, as is the case of ontologies of the tourist domain.

FOCA considers criteria such as the clarity of the ontology, that is the definitions of concepts that arise from social situations. Another criterion is consistency, which guarantees that the ontology is consistent with its purposes. Completeness takes into account the whole meaning of individuals. On the other hand, adaptability refers to the reaction of the ontology to small changes in the axioms, and computational efficiency examines the ease and success by which reasoners can process the ontology.

Concerning the GQM approach, the objectives are defined in questions to extract information from the models. Moreover, the questions define a set of metrics for interpretation. In this way, the FOCA methodology raises five verification objectives. For each objective, a set of questions is posed (thirteen in total) that seek to interpret six metrics.

Regarding the last step of the FOCA methodology, the quality check, the evaluator verifies the questions and calculates their grades using the beta regression models proposed by Ferrari [70]. The authors of FOCA considered this model very appropriate since it is commonly used to model random varieties that assume values in the interval of the unit (0, 1), such as rates, percentages, and proportions. The beta density can show different forms depending on the values of the parameters. Finally, it should be clarified that the authors recognized that there are questions with some degree of subjectivity, especially Questions 7 to 9, which can affect the final score; however, the regression model considers different weights for each of the parameters.

The ontology's quality was calculated by the beta regression models [70], as shown in Equation (1):

$$\begin{aligned}
 x &= -0.44 + 0.03(Cov_S \cdot Sb)_i + 0.02(Cov_C \cdot Co)_i + 0.01(Cov_R \cdot Re)_i \\
 &\quad + 0.02(Cov_{Cp} \cdot Cp)_i - 0.66 \cdot LExp_i - 25(0.1 \cdot NI)_i \\
 \hat{\mu}_i &= \frac{\exp(x)}{1 + \exp(x)}
 \end{aligned} \tag{1}$$

where:

Cov_S = Goal 1 grade;

Cov_C = Goal 2 grade;

Cov_R = Goal 3 grade;

Cov_{Cp} = Goal 4 grade;

$LExp$ = experience of the evaluator; vast experience: $LExp$ is one, if not, zero; NI = one only if some goal was impossible to answer all the questions;

$Sb = 1, Co = 1, Re = 1, Cp = 1$ = because the total quality considers all the roles.

The equation, using the goal grades and considering that the evaluators have some experience, is:

$$\begin{aligned}
 x &= -0.44 + 0.03(83.3 \cdot 1) + 0.02(75.0 \cdot 1) + 0.01(100.0 \cdot 1) + 0.02(75.0 \cdot 1) - 0.66 \cdot 0 - 25(0.1 \cdot 0) \\
 \hat{\mu} &= \frac{\exp(6.059)}{1 + \exp(6.059)} \\
 \hat{\mu} &= 0.9977
 \end{aligned} \tag{2}$$

Thus, the total quality of the ontology is 99% (Equation (2)), which shows that the ontology's quality is high. Thus, OntoTouTra was successfully validated and verified.

5.2. Conceptual Validation

To validate the conceptual model, we used a set of tests applied to a use case to demonstrate the effectiveness of the OntoTouTra ontology using SPARQL queries. These tests were designed with an approach oriented toward the data of real cases gathered from one of the OTAs, using web-scraping techniques. The algorithm was executed with data from Colombia as a tourist destination, which was the selected use case. To answer the questions of the experts [71] in the TTS knowledge domain, we set some KPIs based on [72]. The indicators were grouped into four boxes: Satisfaction, Economy, Sustainability, and Organizational.

The KPIs are interpreted in the knowledge base as questions (CQ) that are answered through queries to the ontology. For each KPI, we developed test cases using SPARQL queries. We chose some KPIs from the document and adapted other indicators according to the TTS. We chose the ten most representative KPIs for the test, taking into account space–time variables in the queries. Furthermore, these queries can be broken down into different levels of grouping and detail, such as geographic areas, timelines, services, tourist experiences, and types of accommodation, among others. In the ten selected queries, we tried to involve these types of groupings in general detail. The selected KPIs are depicted in Table 10.

Table 10. KPI list.

Box	KPI	Indicator
1	01	% of visitors who rate the overall visitor experience as good or excellent
1	02	% of customers who consider the overall impression of the WiFi service to be good or excellent
2	03	Number of day visitors
3	04	Number of tourism enterprises (accommodation) per 10,000 population
3	05	Ratio of number of reviews to local population
3	06	Population rate with hotel influence
2	07	Foreign tourist arrivals (FTAs)
2	08	Inbound and domestic tourism
2	09	Seasonality patterns
2	10	Tourist experiences

5.3. Ontology Testing

The approach to using KPIs as test cases allowed evaluating the ontology from several indicators: semantics, inferences from ontological terms, consistency of the purpose of the ontology, and detection of inconsistencies. Table 10 depicts the test cases for each of the selected KPIs. As a reference for comparison, local government and WTO sources were sought to contrast the expected results (see Table 11). The test cases were run using SPARQL queries whose results demonstrated the reliability of the ontology when compared with the expected results (see the Supplementary Material). The execution of the test cases was performed with the Apache Jena Fuseki tool. The results are evidenced in the Supplementary Material.

In Table 11, we observe the results of the ontology test from a conceptual point of view, according to the application domain. The column “Test case” corresponds to the KPIs to validate. The column “Expected results” corresponds to the projected results after the test case (SPARQL query) has been executed. We compared these results with the sources, which are tourism authorities indicated in the column “Comparison sources.” From these sources, we identified the comparison data shown in the column “Source’s data.” We obtained results when executing the SPARQL queries, and these are listed in the column “Results obtained.” Based on these last two columns, we compared the consistency of the results. This comparison must be considered proportionally. The data from these sources were consolidated from the tourism sector, while the ontology data came from a portion of this sector that we obtained from the OTAs.

Table 11. Expected results.

Test Case	KPI	Expected Results	Comparison Sources	Source's Data	Results Obtained	Note
T001	1	Over 60 % of visitors rated the experience as good or excellent		-	71.56%	
T002	2	In Colombia, over 50% of customers considered the WiFi service to be good or excellent		-	53.5%	
T003	3	In Colombia, in 2019, over 1000 reviews per day	Colombia's Fact Sheets [73] pages 1–2	4,100,000 annual (2019)	2423 (mean)	Booking's reviewers represent the 21.57% visitors
T004	4	In Colombia, two (2) accommodation enterprises per 10,000 population	Colombia's Fact Sheets [73] page 4	5.6	2.33	28,000 establishments/50 million inhabitants = 5.6. Booking = 2.33
T005	5	The number of reviews depends on the local tourism industry (33 departments in Colombia)	[54] page 18	Bogotá, Antioquia, Bolívar	Bogotá, Antioquia, Bolívar	Top-3 departments
T006	6	Population rate with hotel influence depends on the local tourism industry	Colombia's Tourism Report [54] page 28	San Andrés, Bolívar, Bogotá	Bogotá, San Andrés, Valle	Top 3 departments
T007	7	Top 10 foreign tourist arrivals (FTAs) in Colombia	Colombia's Tourism Report [54] page 7	USA, Peru, France	USA, France, Argentina	Top 3 countries
T008	8	Inbound and domestic tourism in Colombia per department	Colombia's Fact Sheets [73] pages 1–2	4,100,000	459,322	Inbound travels
T009	9	Seasonality patterns per month of 2019 in Colombia	UNWTO Seasonality [74]	January–March, July–August	January–April, July–August	Peak seasons
T010	10	Top 10 Tourist experiences in Colombia		-	Beach, tours, game room	Top 3 tourist experiences

The results in Table 11 demonstrate the OntoTouTra ontology’s effectiveness in retrieving conceptual information from the TTS domain. All the proposed indicators were achieved through the SPARQL queries. In addition, the open architecture of this ontology allows the use of different tools and technologies to access data from the endpoint, such as Apache Fuseki, Apache Jena, Protégé, Open Link Virtuoso, Fuseki SOH (REST API), and OBAs. For this reason, the column “Note” describes the special comparison considerations for each test case.

6. Analysis of the Results

The objective of this work was to provide a knowledge base for the tourist traceability system. This knowledge base was built with input data from ubiquitous data, mainly social networks, such as OTAs. This paper indicates the method to construct an ontology whose data sources are typical in Big Data environments. The features of the developed ontology called OntoTouTra are depicted in Table 12. In the Supplementary Material, we show the screenshots running OntoTouTra on each of these tools.

Table 12. OntoTouTra features.

Item	Feature	Tool
1	SPARQL Interface	Apache Jena Apache Jena Fuseki Protégé OpenLink Virtuoso
2	Web interface	RDFLib/Dash WebVOWL/TikZ [75]
3	REST API	Fuseki SOH Ontology-Based API (OBA)
4	Documentation	Protégé OBA

Table 13 summarizes the differences between OntoTouTra and the similar ontologies within the tourism domain, based on the studies of [76,77]. Each ontology has its specific purpose within the field of tourism. For its development, common standards were used to generate the axioms. The number of concepts depends on the domain contemplated. When evaluating the ontology with use cases based on the KPIs, it was a challenge that we overcame when performing complex SPARQL queries, especially in the space–time dimensions that are sensitive in a TTS.

Table 13. OntoTouTra vs. other tourism ontologies.

Item	Domain	Use	Axioms
OntoTouTra	Tourist traceability	Decision-making at the destination	OWL
Mondeca	Tourism	Tourism concepts	OWL
HarmoNET	Tourism	Accommodation	OWL
Travel Itinerary	Travel	Tourist itineraries	OWL
Hontology	Hotel	Hotels	OWL
OnTour Project	e-Tourism	Accommodation	OWL
COTRIN	Open Travel Alliance (OTA) specifications	Travel industry	XML schema
LA_DMS project	DMO	Tourist destination	OWL-S
Hi-Touch project	Tourism products	Customer’s expectations	OWL
TAGA	Travel concepts	Simulations	OWL

In this study, we used FOCA [69] as an ontology evaluation method because it allowed us to evaluate multiple quality criteria, which were the criteria based on Gruber’s proposal [66] and served as the metrics of evaluation. Following FOCA and the beta

regression modeling equation [70], the total quality was calculated based on the weights of each metric of the evaluation goals. In this way, a total quality score was obtained for the OntoTouTra ontology, taking into account the TTS domain, of 99.77%, indicating that the quality was high and satisfied the requirements of its domain. To achieve greater objectivity in this weighting, we used ontology evaluation tools such as HermiT [67] and OOPS! [68]. The first one allowed us to provide the reasoning for the consistency of the content of the ontology, and the second one detected the pitfalls. The results generated by both tools were satisfactory.

7. Data Treatment

This paper presents the methodology of constructing a tourist traceability ontology called OntoTouTra as an educational and research effort. The data to generate the individuals (instances) were obtained from ubiquitous computing sources, especially from social networks, sensors installed in POIs, and applications installed on users' mobile devices. The OntoTouTra ontology, without individuals, and the source code referred to in this paper are available in the repository indicated in Appendix A of this paper. We can run the source code to obtain the data and feed the ontology with the individuals. Still, before doing this, we strongly recommend that the ToSs be reviewed for the data treatment of the owner or owners of these data.

For our case, we reviewed the ToS of Booking.com [78], which was the OTA that we chose to scrape the data to carry out the test cases and the study case. Within these ToS, in the "Scope & Nature of Our Service" Section, we find "... Our Trip Service is made available for personal and non-commercial use only. Therefore, you are not allowed to resell, deep-link, use, copy, monitor (e.g., spider, scrape), display, download, or reproduce any content or information, software, reservations, tickets, products, or services available on our Platform for any commercial or competitive activity or purpose...". On the other hand, in the "Intellectual Property Rights" Section, we find: "... Booking.com exclusively retains ownership of all rights, title and interest in and to (all intellectual property rights of) (the look and feel (including infrastructure) of) the Platform on which the service is made available (including the guest reviews and translated content) and you are not entitled to copy, scrape, (hyper-/deep) link to, publish, promote, market, integrate, utilize, combine or otherwise use the content (including any translations thereof and the guest reviews) or our brand without our express written permission...". We can also consult the "robots.txt" file of the OTA website to verify if it prevents (disallows) crawling or scraping and from the crawl rate to verify if the query is made by a human.

The objective of Krotov and Silva's research [79,80] was to identify a set of ethical and legal considerations when collecting data from the web using automated tools. According to them, no legislation directly addresses web scraping. There is a set of theories and laws that guide web scraping, such as "copyright infringement," "breach of contract" on the side of the web user, the act of computer fraud and abuse (CFAA), and "trespass to chattels." In the case of copyrighted material, data that are explicitly owned and copyrighted by the website owner may lead to a case of "copyright infringement." However, a website does not necessarily own user reviews. Given these conditions, and based on the research reflections, we decided to publish the ontology without the individuals (instances). However, the experimentation environment can be reproducible by feeding this ontology with the data obtained after running the software.

8. Discussion and Conclusions

In this study, we proposed a model for building an ontology of a TTS, answering the research question "How can we develop a tourist traceability ontology based on gathering and processing ubiquitous data, using Big Data techniques?" The gap demonstrated in the state-of-the-art showed us the lack of an ontology whose domain was tourist traceability. Therefore, we proposed a model for the creation of the OntoTouTra ontology. In turn, we adapted the lifecycle of Big Data analytics presented by Erl et al. [16] to deal with the

volume, variety, and velocity of data coming from ubiquitous sources, in particular from an OTA.

We applied the GQM approach of the FOCA methodology to validate the OntoTouTra ontology and achieved a score of 99.77% of the total quality of the ontology. We used HermiT, Protégé, and OOPS! as evaluation tools. However, the number of individuals in the ontology, especially tourist reviews, required enormous computational resources. For instance, we used HermiT as a Protégé [56] plugin, and the capacity of this tool restricted its execution. For the evaluation tests, we had to ignore the individuals of the tourist reviews. A new research challenge arises to adapt this type of ontological tool to Big Data environments. The amount of knowledge affects the quality of the ontology's testing processes, which is imperative in this environment. The analysis of the ontology validation results demonstrated its functionality. The validation was conceptual, whose aim was to evaluate the purpose and functionality of the ontology. This goal was achieved by executing SPARQL queries for 10 KPIs representative of a TTS.

As contributions of this study, we highlight the construction model of the ontology, the adaptation of the lifecycle of Big Data analytics so that the ontology works with ubiquitous data sources in Big Data contexts, and the interoperability of the ontology with open systems, since it allows SPARQL queries and RESTFUL API. The source code allowed the creation, access, and use of the ontology in Big Data environments, using PySpark, and the provision of the ontology for open link data, in particular with GeoNames and Time Ontology. The results of this study are a meaningful contribution to the scientific community and to DMOs looking for a knowledge base to support decision-making regarding destination management.

The practical application of the developed ontology is extensive: it serves as a knowledge base to support decision-making in the destination, recommendation systems for tourist experiences, monitoring of the management of the DMO, the design or improvement of tourist experiences, the benchmarking of tourist experiences, tourist service providers, and web portals on destination tourist information, among others.

Through the OntoTouTra ontology, we plan to consolidate the knowledge base for DMOs. As future work, we will include other ubiquitous computing sources, such as data from tourist mobile devices and sensors from POIs. Besides, we will offer a portfolio of tourist experiences of the destination.

Supplementary Materials: The following are available at www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, J.F.M.-M., L.S.-G., A.F.V. and G.R.-G.; data curation, J.F.M.-M.; formal analysis, J.F.M.-M.; funding acquisition, G.R.-G.; investigation, J.F.M.-M.; methodology, J.F.M.-M. and L.S.-G.; project administration, J.F.M.-M.; software, J.F.M.-M.; supervision, A.F.V. and G.R.-G.; validation, J.F.M.-M., L.S.-G., A.F.V. and G.R.-G.; visualization, J.F.M.-M. and L.S.-G.; writing—original draft, J.F.M.-M.; writing—review and editing, J.F.M.-M. and L.S.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was funded by the Universidad del Cauca (501100005682).

Acknowledgments: This research was financially supported by the Ministry of Science, Technology, and Innovation of Colombia (733-2015) and by the Universidad Santo Tomás Seccional Tunja.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

API	application programming interface
CDA	confirmatory data analysis
CQ	competency questions
DMO	destination management organization
EDA	exploratory data analysis
eWOM	electronic word-of-mouth
GQM	goal-question-metric approach
IoT	Internet of Things
ISO	International Organization for Standardization
KM	knowledge management
KPI	key performance indicator
NLP	natural language processing
OntoTouTra	Ontology for Tourist Traceability
OTA	online travel agency
OWL	Web Ontology Language
POI	point of interest
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SPARQL	SPARQL Protocol and RDF Query Language
ToSs	terms of service
TTS	tourist traceability system
UNWTO	United Nations World Tourism Organization

Appendix A. Ontology Repository

The source code to build the OntoTouTra ontology and obtain its individuals (instances) from an OTA is available in the following public repository, including installation instructions: <https://github.com/jfmendezam/ontotoutra>.

Furthermore, we can find the repository of the ontology and its documentation at <http://tourdata.org/>.

References

- Chantre Astaiza, A.; Fuentes-Moraleda, L.; Muñoz-Mazón, A.; Ramirez-Gonzalez, G. Science Mapping of Tourist Mobility 1980–2019. Technological Advancements in the Collection of the Data for Tourist Traceability. *Sustainability* **2019**, *11*, 4738. [CrossRef]
- Schuitmaker, R.; Xu, X. Product traceability in manufacturing: A technical review. *Procedia CIRP* **2020**, *93*, 700–705. [CrossRef]
- ISO. *ISO 12875:2011. Traceability of Finfish Products*. Available online: <https://www.iso.org/obp/ui/#iso:std:iso:12875:ed-1:v1:en> (accessed on 2 November 2019).
- GS1. *The GS1 Traceability Standard: What You Need to Know*; Technical Report; Global Office: Brussels, Belgium, 2007. Available online: https://www.gs1.org/docs/traceability/GS1_traceability_what_you_need_to_know.pdf (accessed on 2 November 2019).
- Chandrasekaran, B.; Josephson, J.; Benjamins, V.R. What Are Ontologies, and Why Do We Need Them? *IEEE Intell. Syst. Their Appl.* **1999**, *14*, 20–26. [CrossRef]
- Xiang, Z.; Gretzel, U.; Fesenmaier, D. Semantic Representation of Tourism on the Internet. *J. Travel Res.* **2009**, *47*, 440–453. [CrossRef]
- Tribe, J.; Liburd, J.J. The tourism knowledge system. *Ann. Tour. Res.* **2016**, *57*, 44–61. [CrossRef]
- Mouhim, S.; Aoufi, A.; Cherkaoui, C.; Hassan, D.; Mammass, D. A knowledge Management Approach Based on Ontologies: The Case of tourism. *Int. J. Comput. Sci. Emerg. Technol.* **2011**, *2*, 362–369.
- Uschold, M.; Grüninger, M. Ontologies: Principles, methods and applications. *Knowl. Eng. Rev.* **1996**, *11*, 93–136. [CrossRef]
- Missikoff, M.; Taglino, F. An Ontology-based Platform for Semantic Interoperability. In *Handbook on Ontologies*; Staab, S., Studer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 617–633.
- Carloni, O. Boolean Formulas of Simple Conceptual Graphs SGBF. In Proceedings of the Second International Conference on Graph Structures for Knowledge Representation and Reasoning, Barcelona, Spain, 16 July 2011; pp. 18–67.
- Siorpaes, K.; Bachlechner, D. OnTour: Tourism Information Retrieval based on YARS. In Proceedings of the 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, 11–June 2006.
- Prantner, K.; Ding, Y.; Luger, M.; Yan, Z.; Herzog, C. Tourism ontology and semantic management system: State-of-The-Arts analysis. In Proceedings of the IADIS International Conference: IADIS, Vila Real, Portugal, 5–8 October 2007.

14. Siricharoen, W.V. Using Ontologies for E-tourism. In Proceedings of the 4th WSEAS/IASME International Conference on Engineering Education (EE 2007) Proceeding, Crete Island, Greece, 24–26 July 2007.
15. Zhao, X.; Liu, L.; Wang, H.; Song, W. Ontology Construction of the Field of Tourism in Africa. In Proceedings of the 2015 8th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 12–13 December 2015; pp. 47–50. [[CrossRef](#)]
16. Erl, T.; Khattak, W.; Buhler, P. *Big Data Fundamentals: Concepts, Drivers & Techniques*; ServiceTech Press: Englewood Cliffs, NJ, USA, 2016.
17. Huang, Y.; Bian, L. Using Ontologies and Formal Concept Analysis to Integrate Heterogeneous Tourism Information. *IEEE Trans. Emerg. Top. Comput.* **2015**, *3*, 172–184. [[CrossRef](#)]
18. Valls, A.; Gibert, K.; Orellana, A.; Antón-Clavé, S. Using ontology-based clustering to understand the push and pull factors for British tourists visiting a Mediterranean coastal destination. *Inf. Manag.* **2018**, *55*, 145–159. [[CrossRef](#)]
19. Miller, G.; Beckwith, R.; Fellbaum, C.; Gross, D.; Miller, K. Introduction to WordNet: An On-line Lexical Database. *Int. J. Lexicogr.* **1991**, *3*, 235–244. [[CrossRef](#)]
20. Islam, M.R.; Hossain, B.A.; Imteaj, M.N.; Akhter, S.; Jogesh, H.S.; Mostafa, M.B. OnTraNetBD: A knowledgebase for the travel network in bangladesh. In Proceedings of the 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, 21–23 December 2017; pp. 170–174.
21. Giunchiglia, F.; Dutta, B. DERA: A Faceted Knowledge Organization Framework. In Proceedings of the International Conference on Theory and Practice of Digital Libraries, Lyon, France, 25–27 August 2011.
22. Suchanek, F.; Kasneci, G.; Weikum, G. Yago: A Large Ontology from Wikipedia and WordNet. *J. Web Semant.* **2008**, *6*, 203–217. [[CrossRef](#)]
23. Rodríguez-García, M.; Valencia-García, R.; Garcia-Sanchez, F.; Samper Zapater, J.J. Creating a semantically-enhanced cloud services environment through ontology evolution. *Future Gener. Comput. Syst.* **2014**, *32*, 295–306. [[CrossRef](#)]
24. Llorens, J.; Morato, J.; Génova, G.; Fuentes, J.; Quintana, V.; Díaz, I. RHSP: An Information Representation Model Based on Relationship. *Stud. Fuzziness Soft Comput.* **2004**, *159*, 221–253.
25. Santamaria-Granados, L.; Mendoza-Moreno, J.F.; Ramirez-Gonzalez, G. Tourist Recommender Systems Based on Emotion Recognition—A Scientometric Review. *Future Internet* **2021**, *13*, 2. [[CrossRef](#)]
26. Chu, Y.; Wang, H.; Zheng, L.; Wang, Z.; Tan, K.L. TRSO: A Tourism Recommender System Based on Ontology. In Proceedings of the International Conference on Knowledge Science, Engineering and Management, Passau, Germany, 5–7 October 2016; Volume 9983.
27. Guergour, H.E.; Boufaïda, Z. A domain ontology building process based on principles of social web. In Proceedings of the 2012 International Conference on Information Technology and e-Services, Las Vegas, NV, USA, 16–18 April 2012; pp. 1–6.
28. Moreno, A.; Valls, A.; Isern, D.; Marin, L.; Borràs, J. SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities. *Eng. Appl. Artif. Intell.* **2013**, *26*, 633–651. [[CrossRef](#)]
29. Shoal, N.; Ahas, R. The use of tracking technologies in tourism research: The first decade. *Tour. Geogr.* **2016**, *18*, 587–606. [[CrossRef](#)]
30. Girardin, F.; Calabrese, F.; Dal Fiore, F.; Ratti, C.; Blat, J. Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive Comput.* **2009**, *7*, 36–43. [[CrossRef](#)]
31. Mariani, M.; Borghi, M. Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tour. Manag.* **2018**, *66*, 47–52. [[CrossRef](#)]
32. Lytvyn, V.; Vysotska, V.; Burov, Y.; Demchuk, A. Architectural Ontology Designed for Intellectual Analysis of E-Tourism Resources. In Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 11–14 September 2018; Volume 1, pp. 335–338.
33. Lee, C.I.; Hsia, T.C.; Hsu, H.C.; Lin, J.Y. Ontology-based tourism recommendation system. In Proceedings of the 2017 4th International Conference on Industrial Engineering and Applications (ICIEA), Nagoya, Japan, 27–29 April 2017; pp. 376–379.
34. Smirnov, A.; Ponomarev, A.; Shilov, N.; Kashevnik, A.; Teslya, N. Ontology-Based Human-Computer Cloud for Decision Support: Architecture and Applications in Tourism. *Int. J. Embed. Real-Time Commun. Syst.* **2018**, *9*, 1–19. [[CrossRef](#)]
35. Prasamarso Kuntarto, G.; Gunawan, I.; Moechtar, F.; Ahmadin, Y.; Santoso, B.I. Dwipa Ontology III: Implementation of Ontology Method Enrichment on Tourism Domain. *Int. J. Smart Sens. Intell. Syst.* **2017**, *10*, 903–919.
36. Borràs, J.; Flor, J.; Perez, Y.; Moreno, A.; Valls, A.; Isern, D.; Orellana, A.; Russo, A.; Clavé, S. SigTur/E-Destination: A System for the Management of Complex Tourist Regions. In *Information and Communication Technologies in Tourism*; Springer: Vienna, Austria, 2011; pp. 39–50.
37. Wick, M. *GeoNames Ontology*; Technical Report; Unxos GmbH: Wollerau, Switzerland, 2015. Available online: <http://download.geonames.org/export/dump/readme.txt> (accessed on 21 March 2019).
38. Frontini, F.; Del Gratta, R.; Monachini, M. GeoDomainWordNet: Linking the GeoNames Ontology to WordNet. In Proceedings of the Language and Technology Conference, Poznań, Poland, 7–9 December 2016; Volume 9561, pp. 229–242.
39. Team, G. GeoNames Webservice Subdivision Levels. Available online: <https://www.GeoNames.org/export/subdiv-level.html> (accessed on 21 March 2019).
40. DANE. Geovisor de Consulta de Codificación de la Divipola. Available online: <https://geoportal.dane.gov.co/geovisores/territorio/consulta-divipola-division-politico-administrativa-de-colombia/> (accessed on 21 March 2019).

41. Cox, S.; Little, C. *Time Ontology in Owl*. Available online: <https://www.w3.org/TR/owl-time/> (accessed on 21 March 2019).
42. International Open Data Charter ODC. *ODC Principles*. Available online: <https://opendatacharter.net/adopt-the-charter/> (accessed on 21 March 2019).
43. Ministerio de Tecnologías de la Información y las Comunicaciones. Datos Abiertos. Available online: <https://www.datos.gov.co/> (accessed on 21 March 2019).
44. Situr Boyacá. Sistema de Información Turística de Boyacá. Available online: <https://situr.boyaca.gov.co/> (accessed on 21 March 2019).
45. Lohmann, S.; Negru, S.; Haag, F.; Ertl, T. Visualizing Ontologies with VOWL. *Semant. Web* **2016**, *7*, 399–419. [CrossRef]
46. Fernández-López, M.; Gomez-Perez, A.; Juristo, N. METHONTOLOGY: From ontological art towards ontological engineering. In Proceedings of the Engineering Workshop on Ontological Engineering (AAAI97), Stanford, CA, USA, 24–26 March 1997.
47. Kumara, B.; Paik, I.; Zhang, J.; Siriweera, T.H.A.; Koswatte, K. Ontology-Based Workflow Generation for Intelligent Big Data Analytics. In Proceedings of the Conference: IEEE International Conference on Web Services (ICWS 2015), New York, NY, USA, 27 June–2 July 2015. [CrossRef]
48. Booking. Booking.com Home Page. Available online: <https://www.booking.com/> (accessed on 9 April 2019).
49. Expedia. Expedia.com Home Page. Available online: <https://www.expedia.com/> (accessed on 9 April 2019).
50. Airbnb. Airbnb.com Home Page. Available online: <https://www.airbnb.com/> (accessed on 9 April 2019).
51. TripAdvisor. TripAdvisor.com Home Page. Available online: <https://www.tripadvisor.com/> (accessed on 9 April 2019).
52. MinCIT. Prestadores Registro Nacional de Turismo—Datos Abiertos. Available online: <https://www.datos.gov.co/Comercio-Industria-y-Turismo/Prestadores-Registro-Nacional-de-Turismo/npkw-6rke> (accessed on 21 March 2019).
53. Bermudez, Y.; Aponte, A.; Zuluaga, V.; Moreno, C.; Ceballos, O. Prototipo de Publicación de Datos Turísticos Apoyados en Linked Open Data Para el Consumo de Información del Sector Ecoturístico en el Centro del Valle del Cauca. Available online: <https://bibliotecadigital.univalle.edu.co/handle/10893/14492> (accessed on 21 March 2019).
54. Ministerio de Comercio, Industria y Turismo. Informes de Turismo. Available online: <https://www.mincit.gov.co/estudios-economicos/estadisticas-e-informes/informes-de-turismo> (accessed on 21 March 2019).
55. Osorio, M.; Garijo, D. Ontology-Based APIs (OBA). Available online: <https://oba.readthedocs.io/en/latest/> (accessed on 17 September 2020).
56. Musen, M. The Protégé Project: A Look Back and a Look Forward. *AI Matters* **2015**, *1*, 4–12. [CrossRef] [PubMed]
57. Hardi, J. Cellfie Plugin. Available online: <https://github.com/protegeproject/cellfie-plugin> (accessed on 11 October 2019).
58. Gomez-Perez, A.; Fernández-López, M.; Corcho, O. *Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web*; Springer Science & Business Media: New York, NY, USA, 2004.
59. Steiner, C.; Albert, D. Validating domain ontologies: A methodology exemplified for concept maps. *Cogent Educ.* **2017**, *4*, 1263006. [CrossRef]
60. Glimm, B.; Horrocks, I.; Motik, B.; Stoilos, G.; Wang, Z. HermiT: An OWL 2 Reasoner. *J. Autom. Reason.* **2014**, *53*, 245–269. [CrossRef]
61. Loshin, D. *Big Data Analytics*; Morgan Kaufmann: Amsterdam, The Netherlands, 2013.
62. Bornhorst, T.; Ritchie, J.; Sheehan, L. Determinants of Tourism Success for DMOs & Destinations: An Empirical Examination of Stakeholders’ Perspectives. *Tour. Manag.* **2010**, *31*, 572–589.
63. Emani, C.; Cullot, N.; Nicolle, C. Understandable Big Data: A survey. *Comput. Sci. Rev.* **2015**, *17*, 70–81. [CrossRef]
64. Ceravolo, P.; Azzini, A.; Angelini, M.; Catarci, T.; Cudre-Mauroux, P.; Damiani, E.; Mazak, A.; Van Keulen, M.; Jarrar, M.; Santucci, G.; et al. Big Data Semantics. *J. Data Semant.* **2018**, *7*, 65–85. [CrossRef]
65. Lytvyn, V.; Vysotska, V.; Veres, O.; Brodyak, O.; Oryshchyn, O. Big Data analytics ontology. *Technol. Audit. Prod. Reserv.* **2017**, *1*, 16–27. [CrossRef]
66. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum. Comput. Stud.* **1995**, *43*, 907–928. [CrossRef]
67. Classification, O.O. Birte Glimm and Ian Horrocks and Boris Motik and Giorgos Stoilos. In Proceedings of the 9th International Semantic Web Conference (ISWC 2010), Shanghai, China, 7–11 November 2010; Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B., Eds.; Springer: Shanghai, China, 2010; Volume 6496, pp. 225–240.
68. Poveda-Villalón, M.; Gomez-Perez, A.; Suárez-Figueroa, M.C. OOPS! (Ontology Pitfall Scanner!): An on-line tool for ontology evaluation. *Int. J. Semant. Web Inf. Syst.* **2014**, *10*, 7–34. [CrossRef]
69. Bandeira, J.; Bittencourt, I.; Espinheira, P.; Isotani, S. FOCA: A Methodology for Ontology Evaluation. *arXiv* **2016**, arXiv:1612.03353.
70. Ferrari, S.; Cribari-Neto, F. Beta Regression for Modelling Rates and Proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [CrossRef]
71. Bezerra, C.; Freitas, F.; da Silva Santana, F. Evaluating Ontologies with Competency Questions. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, 17–20 November 2013; pp. 284–285. [CrossRef]
72. Office for National Statistics. *Measuring Tourism Locally*; ONS: Newport, UK, 2010.
73. UNWTO. Country Fact Sheets—Colombia. Available online: <https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2020-10/colombia.pdf> (accessed on 4 February 2020).
74. UNWTO. Tourism Seasonality across Destinations. Available online: <https://www.unwto.org/seasonality> (accessed on 4 February 2020).

75. Tantau, T. *The TikZ and PGF Packages—Manual for Version 3.1.9a*; Institut für Theoretische Informatik, Universität zu Lübeck: Lübeck, Germany, 2021.
76. Chaves, M.; Trojahn, C. Towards a Multilingual Ontology for Ontology-driven Content Mining in Social Web Sites. In Proceedings of the ISWC 2010 Workshops, Shanghai, China, 7–8 November 2010.
77. Sicilia, M.A. *Handbook of Metadata, Semantics and Ontologies*; World Scientific: Singapore, 2013; pp. 393–406. [[CrossRef](#)]
78. Booking. Trip Terms and Conditions. Available online: <https://www.booking.com/content/terms.html> (accessed on 9 April 2019).
79. Krotov, V.; Silva, L. Legality and Ethics of Web Scraping. In Proceedings of the Twenty-Fourth Americas Conference on Information Systems, New Orleans, LA, USA, 16–18 August 2018.
80. Mahto, D.K.; Singh, L. A dive into Web Scraper world. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 689–693.