

HONTOLOGY: A MULTILINGUAL ONTOLOGY FOR THE ACCOMMODATION SECTOR IN THE TOURISM INDUSTRY

Marcirio Silveira Chaves¹ and Larissa Freitas² and Renata Vieira²

¹*Business and Information Technology Research Center (BITREC), Universidade Atlântica, Lisbon, Portugal*

²*Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil*
marcirioc@uatlantica.pt, larissa.freitas@acad.pucrs.br, renata.vieira@pucrs.br

Keywords: Knowledge Engineering, Ontology Development, Multilingual Ontology, Social Semantic Web, Accommodation Sector

Abstract: Ontologies have been used to support both web agents reasoning and human decision making. However, ontology development is a new area and for some knowledge domains they are still rare. Although ontologies had been developed in the context of Semantic Web, it is the Web 2.0 content that is actually pervasive on the web. One of the properties of this content is to be multilingual, which requires multilingual resources to deal with it. Online reviews are examples of multilingual texts provided by products and services consumers. This paper presents a multilingual ontology for the accommodation sector. As a result, we deliver Hontology, a freely available domain-specific ontology. Hontology reuses concepts of other vocabularies such as Dbpedia.org and Schema.org. It is useful for a wide range of applications within the accommodation sector, including ontology-based information extraction, text annotation and information visualisation.

1 INTRODUCTION

Multilingual Semantic Web is one of the most recent challenges to applications dealing with web data, and has gained more attention in the last years (Buitelaar et al., 2010; Montiel-Ponsoda et al., 2011).

Ontologies have been adopted in order to put in practice the vision of Semantic Web. Although the knowledge engineering hard work and time consuming tasks, the building of ontologies for a narrow domain of interest enables the development of semantic-oriented applications. In the tourism industry, the accommodation sector still lacks this kind of resource to support agent reasoning. This sector contains sub-domains such as hotel, hostels and apartments with specific features which make the development of an ontology much more complex.

Ontologies of the accommodations domain can be used to support managerial decision making and end-user applications. Instead of a few categories provided by Web 2.0 portals, e.g. TripAdvisor and Booking.com, accommodation managers are more able to find specific information using an ontology. From the user point of view, it is also easier to click on the concept *indoor swimming pool* and read all the comments mentioning it than to comb hundreds of comments to find opinions about *indoor swimming pool*. Unfortu-

nately, this last scenery probably is the current reality in the most of Small and Medium Accommodations in Europe.

The research presented in this paper has been conducted of a framework to customer management. This framework receives as input a set of online reviews, uses Hontology to annotate them (Chaves et al., 2012b) and presents this information in diverse graphical formats to the decision making (Carvalho and Chaves, 2012). To the best of our knowledge, a robust, coherent and multilingual representation of accommodation sector is lacking to bring into practice the Semantic Web. The main purpose of this paper is filling the gap of literature in describing the process of building a multilingual ontology for the accommodation sector. As a result, we deliver Hontology (H stands for hotel, hostel and hostel), a freely available domain-specific ontology in four languages: English, Portuguese, Spanish and French.

The rest of this paper is structured as follows: Section 2 describes the related works. Section 3 describes the actors using Hontology and presents its requirements through Unified Modelling Language (UML) use cases. Section 4 details the process of ontology development and gives the statistics about Hontology. Section 5 closes the paper with the final remarks.

2 RELATED WORK

Accommodation sector belongs to the tourism or travel domain for which there are some ontologies developed, e.g. Mondeca (mondeca.com) and Harmonet (harmonet.org). However, the vocabulary of available tourism ontologies covers a limited set of concepts often describing the domain from different perspectives due to the restricted application scope from which the ontologies have been elicited (Barta et al., 2009). Barta et al. (2009) produced the cDOTT ontology, a common ontology for the tourism industry in order to support the interoperability of the agents in low-level operations.

Regarding multilingual ontologies, the Monnet project (monnet-project.eu) - Multilingual Ontologies for Networked Knowledge has presented some efforts to put in practice the vision of Multilingual Semantic Web through of the production of ontologies to the hydrographical domain (Aguado-De-Cea et al., 2010). Aguado-De-Cea et al. (2010) use the labeling system supported by RDF(S) and OWL to include multilingual linguistic information in the first version of the ontology. Ou et al. (2008) developed QALL-ME, an ontology for the tourism domain, which was linked to the multilingual terms in EuroWordNet (<http://www.illc.uva.nl/EuroWordNet/>).

In Salim et al. (2010) an ontology for Islamic Portal can retrieve information in three different languages (English, Malay, and Arabic). The framework proposed to help users find the information that they need without worrying about the language, no matter which language is used. The keywords of several languages are mapped to the same concept in an ontology and are therefore given the same meaning.

Multilingual ontologies can also be used to structure an image database. Therefore, the problems related to the ambiguity could be solved. Popescu (2007) developed a multilingual OWL ontology including knowledge in English, Italian, and Spanish which allow real-time processing when a user queries for images.

Other example of application involving multilingual ontologies is Jung (2011), who proposes a multi-agent system for building indirect alignment between multilingual ontologies. This system was tested with multilingual ontologies written in English, Korean, and Swedish languages. A case study was realized in tourism business domain.

3 HONTOLOGY REQUIREMENTS

Hontology was initially created considering a scenario of applications processing reviews about accommodations on Web sites 2.0 and decision support systems (DSS). The conception of an ontology should also take into account the users point of view, which in case of Hontology includes at least six types of users:

Database Administrator: is interested in Hontology in order to better define the database schema to store instances or individuals of Hontology (e.g. reviews, features, polarity and intensity of the reviews). In Description Logic (Baader et al., 2007), the set of these instances is denominated A-Box, while the schema is called T-Box.

Application Developer: is the person who uses the content of Hontology to develop applications that support decision making.

Prospect: is a user seeking for accommodations. She is interested in common features of an accommodation such as cleaning and staff. These features allow her to make comparisons among accommodations based on her specific needs.

Guest: is a former or current customer of an accommodation. Further the same interest of a prospect, a guest can be interested in providing feedback.

Domain Expert: is the person responsible to maintain the concepts and relationships in the ontology. Her tasks include the matching of concepts to existing ontologies and augmenting Hontology according to the new requirements.

Accommodation Manager: includes managers in strategic and managerial levels of the organization such as the Owner of the Hotel, Marketing managers and Service managers.

4 KNOWLEDGE ENGINEERING

Hontology was developed in seven phases as described in Chaves and Trojahn (2010).

1. Identify existing ontologies on related domains;
2. Select the main concepts and properties;
3. Organize concepts and properties hierarchically into categories;
4. Manually translation of the ontology introducing the labels in Portuguese, Spanish and French;
5. Expand concepts and properties based on online reviews manually evaluated;

6. Translate the new concepts and properties;
7. Export the ontology in several formats.

Domain experts updated Hontology, which includes the mapping of the concepts to existing ontologies. Classifications of accommodations vary from country to country and in some continents it is possible to find specific types of accommodations. One of the guidelines followed in the development of Hontology was that proposed in EC (2009). Chaves et al. (2012b) describe an analysis of the reviews about Small and Medium Hotels in Portugal. This analysis was used as input to build the extension of Hontology, which is detailed in the next sections.

4.1 Main Concepts

The definition of the top-level concepts was based on the needs of the main actor. From these concepts, the actors have a broad and depth vision of the domain, which they are looking for information.

Accommodation: This is the substitute concept of *Hospitality* which was the concept used in the first version of Hontology. Instead of extend Hontology to cover restaurants, transportation and event planning, we extend the set of subclasses to specify the different types of Accommodation. The design of the concept *Accommodation* is based on the categories provided by EC (2009). Actually, Accommodation contains sixteen subconcepts including six kinds of hotel under the concept *Hotel*: *Bunker*, *Capsule*, *Cave*, *Ice*, *Tree House*, and *Under Water*. This list can be extended depending on the applications needs. All the sixteen subconcepts of *Accommodation* are listed in the first column of Table 1.

Facility: This concept was split in six subconcepts: *External Facility*, *Internal Facility*, *Room Facility*, *Bathroom Facility*, *Driver* and *Wheel Chair Accessible*. The last two ones could be into *Internal Facility*, but we decide to create a separate subconcept in order to facilitate the navigation. It is worth remembering that *Internal Facility* contains thirty subconcepts.

Room: contains subconcepts for the main accommodation types, that is, *Apartment Room*, *Hostel Room* and *Hotel Room* as depicted in Figure 1. *Hotel Room* contains nine subconcepts including *Double Room*, *Suite Room* and *Family Room*. *Family Room* is still subdivided into three concepts: *Family Junior Suite with Sea View*, *Family Room with Balcony* and *Family Suite with Sea View*. Some special cases in the knowledge engineering evolve typical properties that are repre-

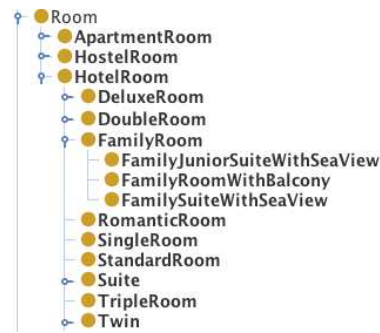


Figure 1: Representation of *Room* types in Hontology.



Figure 2: Representation of *Service* and *Staff* in Hontology.

sented as concepts in Hontology, e.g. with *Balcony* and with *Sea View*.

Service/Staff: Figure 2 presents the subconcepts of the concepts *Service* and *Staff*. Most of the service offered by an accommodation owns a responsible to perform it that is why we decide to present these concepts together.

Guest Type: Figure 3 depicts the main kinds of *Guest Type*. These types are commonly used in travel 2.0 sites in order to classify guests according to a pre-defined profile.

Other concepts to mention include *Design*, *Meal*, *Points of Interest*, *Price*, *Rating* and *Staff.Price* usually is modelled as a property in ontologies, but some of the uses of Hontology may require that price must be explicit as a concept given its relevance to the application. For instance, hotel managers can be interested in reviews mentioning prices of the specific services, such as *BarPrice*, *BreakfastPrice*, *CoffeePrice*, *CotPrice*, *InternetPrice*, *ParkingPrice*, *Restau-*

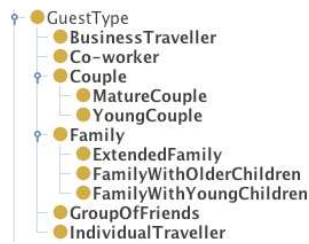


Figure 3: Representation of *Guest Type* in Hontology.

rantPrice, and *RoomPrice*. The same principle was applied to the feature *Timetable*, which is represented as a concept with four subconcepts: *Indoor Swimming Pool Timetable*, *Outdoor Swimming Pool Timetable*, *Restaurant Timetable*, and *Spa Centre Timetable*.

4.2 Exploring Ontology Constructs

This section describes the main constructs used to represent the multilingual and logical aspects in the accommodation sector.

***rdfs:label*, *skos:prefLabel*:** Whenever Hontology is exported to RDF(S) or SKOS, multilingual information is codified using the constructs *rdfs:label* and *skos:prefLabel*, respectively. Into these constructs, we use the attribute *xml:lang* with the respective language as a value. The construct *skos:prefLabel* stores the preferred term in a language.

***owl:equivalentClass*, *owl:sameAs*, *rdfs:subClassOf*:** These labels, at conceptual level, can help the conceptual cross-lingual mappings. That is, concepts from different ontologies described in different languages can be semantically related by using ontology constructs, either to represent taxonomical relations, e.g. *owl:equivalentClass*, *owl:sameAs*, *rdfs:subClassOf* (Gracia et al., 2011). In the version of Hontology described in this paper, we use only the *owl:equivalentClass*, *owl:sameAs* constructs to map concepts to other ontologies (see details in Section 4.4). We do not deal with hierarchical mappings here.

***owl:disjointWith*:** Hontology also handles complex classes through the *owl:disjointWith* construct. For example, all concepts under the concepts *Accommodation*, *GuestType*, *Room*, *Meal*, *PointsOfInterest*, *Price*, *Service*, *Staff* and *Timetable* are codified as *disjoint*.

4.3 Multilingual Aspects

In addition to provide a new ontology for the accommodation sector, we also make an effort to handle

multilingual concepts in order to foster the Multilingual Semantic Web. We start from four languages, i.e. English, Portuguese, Spanish and French. In addition to the main concepts in each language about the accommodation domain, we also store variants whenever they exist. For Portuguese, for instance, we store the variants of European and Brazilian terms. Examples include the concept *Playground* which is called *Praça de Brinquedos* in Brazil and *Parque Infantil* in the European Portuguese, and *Address* which is called *Endereço* in Brazil and *Morada* in Portugal.

Although, XML does not support the explicit markup of variants, applications using Hontology can find this information as a value of the attribute *xml:lang*. This information separately facilitates the storage in linguistic repositories, such as that proposed in Montiel-Ponsoda et al. (2008). We intend to extend the use of variants in other languages than Portuguese, according to the requirements of the applications using Hontology.

4.4 Knowledge Reuse

One of the principles followed along the development of Hontology was the reuse of the existing resources. Although ontologies in the specific domain of accommodations are rare or even inexistent, we can find concepts of this domain formalised in broader domains such as tourism.

The several actors interested in the accommodation information share common concepts that need to be represented or formalised in a single structure or aligned with existing vocabularies. We considered concepts from other public ontologies QALL-ME (Ou et al., 2008), Schema.org and Dbpedia.org. We then aligned our ontology concepts with these other public resources according to Table 1. Moreover, the subconcepts of the concept *Accommodation* in Hontology also include: *Apartment*, *Botel* (*Boat Hotel*), *BungalowAndCamping*, *GuestHouse*, *HolidaySettlement*, *Hostal*, *Inn*, *Pension*, and *OtherAccommodation*.

Table 2 presents the mappings between the concepts of Hontology and QALL-ME. Figure 4 depicts the representation of the concepts *Facility* and *RoomFacility* in both Hontology and QALL-ME ontologies. Hontology maps the concept *Facility* to the existing one in QALL-ME and extends its subconcepts. *BathroomFacility* owns 10 subconcepts, *ExternalFacility* contains nine subconcepts, *InternalFacility* has 29 subconcepts and *RoomFacility* owns 22 subconcepts. It is worth mentioning that QALL-ME was designed to cover the wide domain of tourism and Hontology was built to the specific sector of accom-

Table 1: Alignment between the subconcepts of *Accommodation* in Hontology and QALL-ME, *Lodging Business* in Schema.org and *Tourist Accommodations* in Dbpedia.org.

Hontology	QALL-ME	Schema.org	Dbpedia.org
BedAndBreakfast	BedAndBreakfast	BedAndBreakfast	BedAndBreakfast
Hostel	Hostel	Hostel	Hostel
Hotel	Hotel	Hotel	Hotel
Motel		Motel	Motel
Chalet	Chalet		
Cottage	Cottage		
Resort	Resort		Resort

Table 2: Alignment between the concepts of Hontology and QALL-ME.

Hontology	QALL-ME
Accommodation	Accommodation
Airport	Airport
BusStation	BusStation
Country	Country
Facility	Facility
Location	Location
MetroStation	MetroStation
Price	Price
Restaurant	Restaurant
RestaurantPrice	GastroPrice
RoomPrice	RoomPrice
Stadium	Stadium
Theatre	Theatre
TrainStation	TrainStation

modations. For this reason, Hontology specifies the subconcepts of *Facility* in more detail.

Moreover, it is important to note that the concept *RoomFacility* occurs in both ontologies but it is not aligned. The subconcepts in QALL-ME represent types of room facility while the subconcepts in Hontology represent the infrastructure that is present or in lack in a room.

4.5 Hontology Metrics

Hontology contains 282 concepts categorized into 16 top-level concepts. The concept hierarchy has a maximum depth of 5. Table 3 presents some metrics about Hontology. It is worth mentioning that all concepts and properties are defined in English, Portuguese, Spanish and French. Along the development of this ontology, we observe that Hontology contains the following qualities: i) indefinite expandability, since it remains consistent with increasing content. ii) content and context independence, since any kind of concept can find its place and iii) specify different levels of granularity (levels range from 1 to 5 – see an example in Figure 1).



Figure 4: Representation of the concept *Facility* in Hontology (left) and in QALL-ME (right).

5 FINAL REMARKS

This paper introduces a new resource to put in practice the Multilingual Semantic Web. Hontology, a robust multilingual ontology for the accommodation sector, helps to pave the path for reaching this Web with meaning. Considering that Hontology can be useful for a wide range of knowledge-based systems, we make it freely available from <http://ontolp.inf.pucrs.br/Recursos/downloads-Hontology.php>. Preliminary experiments using Hontology were performed for Portuguese in Chaves et al. (2012a). In this sense, we provide a contribution knowledge engineering, ontology development and domain applications. Future works include the task of ontology population, where we plan to implement algorithms to recognise concepts of the Hontology in the reviews and make them the appropriate instances to the concepts.

Table 3: Hontology Metrics.

Metrics	Values
Number of Concepts	282
Number of Object Properties	8
Number of Data Properties	31
Concept Axioms	
Subconcept axioms	272
Equivalent concepts axioms	4
Disjoint concepts axioms	93
Object Property Axioms	
Functional object property axioms	5
Object property domain axioms	9
Object property range axioms	9
Data Property Axioms	
Functional data property axioms	12
Object data domain axioms	17
Object data range axioms	1

6 ACKNOWLEDGMENTS

We thank Elizabeth Carvalho and Cássia Trojahn for helping in the Spanish and French ontology translation, respectively. We thank the brazilian funding agency CAPES for the scholarship granted.

References

- Aguado-De-Cea, G., Gómez-Pérez, A., Montiel-Ponsoda, E., and Vilches-Blázquez, L. M. (2010). Rivièrè or fleuve? modelling multilinguality in the hydrographical domain. In *MultiLingual Semantic Web (MSW 2010)*.
- Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F., editors (2007). *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2nd edition.
- Barta, R., Feilmayr, C., Pröll, B., Grün, C., and Werthner, H. (2009). Covering the semantic space of tourism: an approach based on modularized ontologies. In *Proc. of the 1st Workshop on Context, Information and Ontologies*, pages 1:1–1:8, New York, NY, USA. ACM.
- Buitelaar, P., Cimiano, P., and Montiel-Ponsoda, E., editors (2010). *1st International Workshop on the Multilingual Semantic Web*. CEUR-WS.org.
- Carvalho, E. and Chaves, M. S. (2012). Exploring user-generated data visualization in the accommodation sector. In *16th International Conference Information Visualisation*, Montpellier, France, 10-13 July. IEEE.
- Chaves, M. S., Freitas, L., Souza, M., and Vieira, R. (2012a). PIRPO: An algorithm to deal with polarity in Portuguese online reviews from the accommodation sector. In *17th International conference on Applications of Natural Language Processing to Information Systems (NLDB)*, Groningen, The Netherlands, 26-28 June.
- Chaves, M. S., Gomes, R., and Pedron, C. (2012b). Decision making based on web 2.0 data: The small and medium hotel management. In *Proceedings of the 20th European Conference on Information Systems (ECIS)*, Barcelona, Spain, 10-13 June.
- Chaves, M. S. and Trojahn, C. (2010). Towards a multilingual ontology for ontology-driven content mining in social web sites. In *Proceedings of the ISWC 2010 Workshops, Volume I, 1st International Workshop on Cross-Cultural and Cross-Lingual Aspects of the Semantic Web*, Shanghai, China.
- EC (2009). Classification of Hotel Establishment within the EU. Available at http://ec.europa.eu/consumers/ecc/docs/hotel_establishment_classification_EU_en.pdf.
- Gracia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., and McCrae, J. (2011). Challenges for the multilingual web of data. *Journal of Web Semantics: Science, Services and Agents on the WWW*, 11(0).
- Jung, J. J. (2011). Exploiting multi-agent platform for indirect alignment between multilingual ontologies: A case study on tourism business. *Expert Systems with Applications*, 38(5):5774–5780.
- Montiel-Ponsoda, E., Aguado de Cea, G., Gómez-Pérez, A., and Peters, W. (2008). Modelling Multilinguality in Ontologies. In *Coling 2008: Companion volume: Posters*, pages 67–70, Manchester, UK.
- Montiel-Ponsoda, E., McCrae, J., Buitelaar, P., and Cimiano, P., editors (2011). *Proceedings of the 2nd International Workshop on the Multilingual Semantic Web*.
- Ou, S., Pekar, V., Orasan, C., Spurk, C., and Negri, M. (2008). Development and alignment of a domain-specific ontology for question answering. In *Proc. of the LREC, 26 May - 1 June, Marrakech, Morocco*.
- Popescu, A. (2007). Image retrieval using a multilingual ontology. In *Computer-Assisted Information Retrieval (Recherche d'Information et ses Applications) - RIAO 2007, 8th International Conference, Carnegie Mellon University, Pittsburgh, PA, USA, May 30 - June 1, 2007. Proceedings, CD-ROM*.
- Salim, J., Hashim, S., and Aris, A. (2010). A framework for building multilingual ontologies for Islamic portal. In *Proc. of the International Symposium in Information Technology (ITSim 2010)*, volume 3, pages 1302–1307.