Providing Personalized Cultural Heritage Information for the Smart Region - A Proposed methodology

Antonino Lo Bue[1], Alan J. Wecker[2,3], Tsvi Kuflik[3], Alberto Machì[1], Oliviero Stock[4]

[1]CNR-ICAR Palermo, Italy
{lobue, machi}@pa.icar.cnr.it
[2]University of Haifa, Israel
[3]University of Trento, Italy
ajwecker@gmail.com, tsvikak@is.haifa.ac.il
[4]FBK-irst, Trento, Italy
stock@fbk.eu

Abstract. In this paper we present a methodology to provide visitors, in smart regions, additional cultural heritage attractions based on prior museum visits using user models and Linked Open Data. Visitor preferences and behavior are tracked via a museum mobile guide and used to create a visitor model. Semantic models and Linked Open Data support the representation of regional assets as Cultural Objects. The visitor model preferences are exploited using a graph similarity approach in order to identify personalized opportunities for visitors by filtering relevant Cultural Objects.

Keywords: Personalization, User Models, Linked Open Data, Smart Regions

1 Introduction

In this short paper we show a blueprint how semantic models and Linked Open Data (LOD) support the representation of regional assets in order to identify categories of opportunities for visitors based on different personal characteristics determined by previous visits. Having a broad infobase from which to cull possibilities is an arduous task that can benefit from automation. Due to the overwhelming number of possibilities, it is important to personalize the Cultural Heritage (CH) experience. When considering what is requires from a smart, personalized system, it becomes clear that the reasoning process of the system has to focus on identifying opportunities for intervention. When and how to intervene and what information to deliver/service to offer. Having a user model, a context model, and a model of the cultural objects are essential for successful support. These can lead to the interaction of museums and places of cultural heritage to create mega-tourist experience (similar to Verbke and Rekom [6] concept of the "museumpark") which can have a positive market effect for the region.

We describe our methodology: First we use exhibits in a museum (we use Castle Buonconsiglio in the Trentino Region as examples throughout this paper) and tag them using semantic concepts. Then a mobile museum guide is used to track visitors.

Based on this data a user model is developed consisting of characteristics and preferences. We then use a dataset of Cultural Objects using an ontological representation of the domain to cull opportunities. Visitor Preferences are used to filter which Cultural Objects are relevant, and Characteristics are used to determine whether an event or cultural heritage place is desired. Context is used to filter for proximate locations weather conditions, opening times, etc. Again characteristics are used to determine how best to present this information to the visitor.

2 Background

In this section we review two technology areas: User Modeling and personalization in CH, and Linked Open Data and Semantic relatedness.

According to Ardisonno et al[2], for more than 20 years, cultural heritage has been a favored domain for personalization research and as soon as mobile technology appeared, it was adopted for delivering context-aware cultural heritage information both indoors and outdoors. For personalization, a system needs to have a model of its user. A number of approaches are possible: Overlay, Feature-based, Content based, and Collaborative filtering. In this proposed methodology we use an implicit content based approach, where user interests are represented as sets of words occurring in the textual descriptions of items relevant for the user. Visitors have been observed to behave in certain stereotypical movement patterns [11]; patterns such as Butterfly, Grasshopper Ant, and Fish[10]. The use of personality types to tailor software is not new. We use the SLOAN Big 5 characterization as it is standard and much research has been done using it [5]. We focus on two traits we believe are connected to the museum experience: Inquisitiveness, which is a measure of curiosity and Orderliness, which measures thoroughness and the need for structure. Introversion and Extroversion could also play a part in group visits, but is not examined in this research. In addition we posit a connection between movement types and "identity" types proposed by John Falk [4]. Preliminary ideas for the connection of movement patterns to personality types have been proposed [1].

Public agencies collect organize and manage a vast amount of data. Local and European projects aims to deliver data as freely available, reusable and distributed without any restriction, the so call Open Data. As part of these initiatives, tourism and cultural heritage datasets have been published as Open Data. Semantic Web technologies and in particular the Linked (Open) Data paradigm, introduced by Sir Tim Berners-Lee in 2006 [3], are opening new ways for data integration and reuse, creating a method to make data interoperable at a semantic level. Ontologies formally represent knowledge as a set of concepts and their relationships within a domain. RDF[1] and OWL[2] standards enable the formal representation of ontologies as set of triples (subject, predicate, object). Ontologies are used to express vocabularies of Linked Data

1 http://www.w3.org/RDF/
2 http://www.w3.org/2001/sw/wiki/OWL

triples. On top of RDF and OWL, SPARQL Query Language[3] is used to query and retrieve information stored as triples thus allowing and facilitating access to the so called Web of Data. DBpedia[4], can be seen as the ontological version of Wikipedia, its the core of the Linked Open Data cloud.

In the Natural Language Processing area, semantic relatedness between terms or concepts can be computed using two main approaches: (1) defining a topological graph similarity using ontologies and computing the minimal graph distances between terms, (2) using statistical methods and word co-occurrence in a corpus and calculating the correlation between words. “WikiRelate!" [8], measures correlation among terms using a graph based distance measure on the Wikipedia categories. The system uses the inverse path length measure as a distance metric for terms correlation. Leal et al [9] present an approach for computing the semantic relatedness of terms using the knowledge base of DBpedia, based on an algorithm for finding and weighting a collection of paths connecting concept nodes. The implemented algorithm defines the concept of proximity rather than the inverse path length distance as a measure of relatedness among nodes. Our methodology is based on the inverse path length measure but we apply this to a graph of ontology terms extracted from DBpedia and used as annotation for Open Data resources. Moreover, we also take into account the concept introduced by Moore et al. [7], that evaluates paths calculating the number of outgoing links of each node, in order to improve the precision of the algorithm.

3 System

The mobile guide, at each position of interest (POI), presents a list of relevant media assets. The mobile guide system logs: the POI, which assets are chosen how long they viewed the asset, and in general how long did they stay at the point of interest. We collect two types of information, the first in order to determine general personal characteristics and the second in order to determine specific topic interests. In general we use movement styles, to predict user characteristics (such as personality). We use time viewing presentations in order to determine user topic preferences.

In order to characterize the user we make use of his general movement activities. We use the following statistics: 1) NumberOfPOIsVisted (NPV) – number of positions where a person stayed more than 9 seconds as detected and logged by the mobile guide's positioning system. Nine seconds is a number we have used for previous analysis and has provided good results. 2) POIsWherePresentationsSeen (PPS) – the number of positions where the visitor viewed at least one media asset connected to that position as computed from the logs of the mobile guide. 3)NumberOfPresentationSeen (NPS) – the total number of media assets the visitor viewed as computed from the logs of the mobile guide.

[3] http://www.w3.org/TR/sparql11-query/
[4] http://dbpedia.org

Table 1. Connecting the user behavior to personality and Falk types

Behavior	*Personality*	*Falk*	*Formula*
Fish	Non curious – Unorderly	Recharger	$((PPS/NPV <= T_1) \& (NPS/PPS < T_3))$
Ant	Inquisitive – Orderly	Explorer	$(PPS/NPV > T_1) \& (NPS/PPS > T_2)$
Grasshopper	Non curious – Orderly	Professional	$(PPS/NPV > T_1) \& (NPS/PPS < T_2)$
Butterfly	Inquisitive – Unorderly	Exp. Seeker	$(PPS/NPV < T_1) \& (NPS/PPS > T_3)$

3.1 What can we find and match up

The system uses annotated internal and external information about cultural places and events. Internal information is taken from catalogues or websites and is used by the mobile guide app to describe user preferences by storing the relevant topics related to exhibits the user has visited and liked. External information is imported from available Open Data about museums and cultural events and enriched in the domain ontology, using knowledge from the Linked Open Data cloud (DBpedia dataset). Data is stored using a domain ontology for tourism called *eTourism*[5]. The ontology covers methodological and practical aspect of services (hotels, B&B, etc.), cultural objects (museum, cultural places, etc.) and events. It is used as a vocabulary model to map external Open Data into RDF triples validated by the ontology concepts. For the present work we have developed a specific module of the *eTourism* ontology named *Cultural Objects Ontology* (*coo*) that covers (1) properties (such as topic, keywords, geographical information) of museums or events, exploits the semantic identity with LOD/DBpedia concepts (using *owl:sameAs* predicates) and implements (2) user profile types and topics of interests selections.

For each museum source, we extract - as a first step, keywords from exhibits of the Castle Buonconsiglio museum. We exploit the semantic relatedness implementing the graph similarity approach. We annotate keywords - for each description, and we disambiguate them to DBpedia concepts using DBpedia Spotlight APIs[6]. We filter out all the not relevant concepts and we then obtain a bag of concepts (related to cultural heritage) like the following:

{dbpedia[7]:Trentino, dbpedia:Prehistory, dbpedia:Ancient_Rome, dbpedia:Middle_Ages, dbpedia:Hunter-gatherer, dbpedia:Upper_Paleolithic, dbpedia:Bronze_Age}

In DBpedia, each concept is related to a category using the property *dcterms:subject*, then each category is part of a hierarchy structure with nodes connected via *skos:broader* properties. For example the below two DBpedia concepts have as *dcterms:subject* the DBpedia *topic* categories:

1) Last_glacial_period *(dcterms:subject) ->{Climate_history, Glaciology, Holocene, Ice_ages}*
2) Ancient_Rome *(dcterms:subject) ->{Ancient_history, Ancient_Rome, Civilizations}*

---

[5] Currently under development at ICAR-CNR within the framework of the national project Dicet-InMoto-Orchestra, (http://www.progettoinmoto.it).
[6] http://spotlight.dbpedia.org
[7] Prefix for http://dbpedia.org/resource/

For the second step, we extract from the DBPedia SPARQL endpoint, for each concept, the *topic* categories of the DBpedia taxonomy. As result we obtain a wider bag of DBpedia *topic* categories describing each museum exhibit. Using the hierarchical structure of categories is thus possible to discover similarities among concepts that have ancestor categories in common.

As external sources, we take the Open Data set delivered by the Italian Cultural Heritage Minister[8] (MIBAC) and we map these objects using the *coo* ontology; then, for each object, we exploit the same process applied for the internal resources, in order to annotate and extract the corresponding bag of topics. As a result, we obtain a list of information for each MIBAC *Cultural Object* (cultural place or event), as in the following example:

foaf:name = "Memorie della Grande Guerra",
coo:mainCategory = http://dbpedia.org/resource/Category:History
Bag of Concepts (*dcterms:description*) ->
{*1918_disestablishments, Aftermath_of_World_War_I, Austria-Hungary, Austria_articles needing attention, States_and_territories_established_in_1867, Anoxic_waters, Back-arc_basins, Contemporary Italian_history, History_of_Austria-Hungary, History_of_modern_Serbia, Wars_involving_Italy, World_War_I* }

In order to select suitable *Cultural Objects* candidates for the user, we define a metric to measure the semantic distance between the user profile tags and the available cultural objects tags. As a first step, we measure the *shortest path distance* between each of the *m topic* categories in the bag of topics of the user profile and the *coo:mainCategory* topic of the suitable candidates (see table 2), and we reduce candidates cardinality by applying an upper threshold to the distance.

Table 2. Example path between two DBpedia categories

Distance	*Steps*	*Distance*	*Steps*
0	dboc:[9]Ancient_history	4	dboc:Art_history
1	dboc:Periods_and_stages_in_archaeology	5	dboc:Visual_arts
2	dboc:Archaeology	6	dboc:Arts
3	dboc:Conservation_and_restoration		

After this step, we refine the result by calculating (via SPARQL queries on the DBpedia endpoint) the *shortest path* between the user bag of topics (*m)* and the suitable candidates bag of topics (*n)* on the remaining subset of cultural objects. Its important to underline that when computing the distance measure between topic categories we also take into account, for each hop of the shortest path, the number of outgoing links of the node: the more outgoing links a node has (to other DBpedia taxonomy nodes) the less it is specific. Broad connected nodes receive low weights while nodes with less outgoing connection will get higher values. We use each pairwise distance as a component of a normalized vector of distances, we evaluate, for each museum or

8 http://dbunico20.beniculturali.it/DBUnicoManagerWeb/#home
9 Prefix for http://dbpedia.org/resource/Category:

event an average normalized distance for each *m* user category and we sum all these distances to define the relatedness of each cultural object. Again an empirical threshold on distance is applied to retain a limited number of candidates.

3.2 Use of characteristics

Using behavior types we can tailor the amount and presentation of information. For example for ants and butterflies we can give ten items. For grasshoppers and fish we may only give two items. Ants and grasshoppers may be given places while butterflies and fish may be given events. Additional personalization may be possible.

4 Discussion and Conclusion

The results we get for the four sample users are shown on the table below.

Table 3. Simulated output of the system with Places and Events suggested per each user behavior. Suggested items are marked with a *.

Type	*Preferences*	*Places, Events*
Ant	Bronze_Age (.5), Feudalism (.2), Middle Ages (.5), Ancient Egyptian funerary practices (.1), Civilizations (.2)	Museo archeologico dell'Alto Adige (Archeology) (.6), Area archeologica Palazzo Lodron (Archeology) (.6), Museo delle palafitte del Lago di Ledro (History) (.4), Museo locale di Aldino (Etnography) (.2*)
Grass-hopper	Romantic_art (.4), 20th-century Italian_painters (.3), Postmodern_art (.3), Fresco_painting (.3),Rural_culture (.1)	Museo Rudolf Stolz (Arts) (.6), Museo di arte moderna e contemporanea di Trento Rovereto (Arts) (.5), Museion - Museo d'arte moderna e contemporanea (Arts) (.6), Museo della Val Venosta (Anthropology) (.2*)
Butterfly	World_War_I (.4), Civilizations (.4), 1st-century Roman emperors (.2), History_of_Europe (.6), Rural_culture (.2)	Doni Preziosi, Immagini e Oggetti dalle Collezioni Museali (Exhibition/History) (.5), Storie da Trento all'Europa. Mostra documentaria (Exhibition/History) (.5)
Fish	Romantic_art (.4), 20th-century Italian painters (.3), Bronze_Age (.5), Fresco painting (.3), Rural_culture (.1)	Rinascimenti Eccentrici al Castello del Buonconsiglio (Exhibition/Arts) (.7), Apertura Spazio archeologico Sotterraneo del Sas (Opening/Archeology) (.4)

Our current metric of semantic relatedness doesn't take into account whether the user profile bag of topics is representative of a sufficiently broad range of museums categories to cover their cultural preferences. To balance this, when all/most of the user preferences are of the same topic area (e.g. Prehistory), one or more among suggested items could be chosen from a minor topic category, to elicit variation in user interests. Our current research involves, the implementation of the methodology to the Old City and the Tower of David Museum in Jerusalem, and the evaluation of the user model and the semantic suggestions results.

References

1. Antoniou, A., Lepouras, G., Lykourentzou, I. et al.: Connecting Physical Space, Human Personalities and Social Networks.
2. Ardissono, L., Kuflik, T., Petrelli, D.: Personalization in Cultural Heritage: The Road Travelled and the One Ahead. User Modeling and User-Adapted Interaction, 22 (2011) 1-27
3. Berners-Lee, T. (2006). Design issues: Linked data.
4. Falk, J. H.: Identity and the museum visitor experience. Left Coast Press Walnut Creek, CA (2009)
5. Higgins, D. M., Peterson, J. B., Pihl, R. O. et al.: Prefrontal Cognitive Ability, Intelligence, Big Five Personality, and the Prediction of Advanced Academic and Workplace Performance. J. Pers. Soc. Psychol., 93 (2007) 298
6. Jansen-Verbeke, M., & Van Rekom, J.: Scanning Museum Visitors: Urban Tourism Marketing. Ann. Tourism Res., 23 (1996) 364-375
7. Moore, J. L., Steinke, F., & Tresp, V.: A novel metric for information retrieval in semantic networks. ESWC 2011 Workshops 65-79
8. Strube, M., & Ponzetto, S. P.: WikiRelate! Computing semantic relatedness using Wikipedia. AAAI Vol. 6, pp. 1419-1424 (2006).
9. Leal, J. P., Rodrigues, V., & Queirós, R.: Computing semantic relatedness using DBpedia. SLATE 2012, page 3519, 2012.
10. Veron, E., & Levasseur, M.: Ethnographie de l'exposition. Centre Georges Pompidou (1983)
11. Zancanaro, M., Kuflik, T., Boger, Z. et al.: Analyzing Museum Visitors' Behavior Patterns. User Modeling 2007, (2007) 238-246