

LUMO: The LinkedTV User Model Ontology

Dorothea Tsatsou and Vasileios Mezaris

Information Technologies Institute, Centre for Research & Technology Hellas
{dorothea,bmezaris}@iti.gr

Abstract. This paper introduces the LinkedTV User Model Ontology (LUMO), developed within the LinkedTV EU project. LUMO aims to semantically represent user-pertinent information in the networked media domain and to enable personalization and contextualization of concepts and content via semantic reasoning. The design principles of LUMO and its connection to relevant ontologies and known vocabularies is described.

1 Introduction

Networked Media, i.e. the convergence of traditional media interfaces with Web content, is the focal point of the Television Linked To The Web (LinkedTV)¹ EU project. However, the convergence of voluminous multimedia content avails to a traditional problem of the Web: information overload. Personalization is a key factor that can alleviate this problem. A holistic vocabulary to classify user-pertinent information under would enable efficient elicitation of user preferences and context. An ontology, comprising of a shared, explicit model of domain conceptualizations and their relations can be used as the reference vocabulary for predictive inferencing of user preferences and for targeted content recommendations.

The use of Linked Open Data (LOD) [1] vocabularies for this purpose, like DBPedia [7], YAGO [8], schema.org², is still hampered by the large volume of data [5] and the shallow, malformed structure, as well as the lack of important user-pertinent semantics or inclusion of semantics not relevant to a user [9]. This poses scalability, coverage and inconsistency problems, while the data volume restricts their use to the server-side, thus compromising user privacy. Conversely, as also argued in [9], upper formal ontologies consist of too abstract conceptualisations to meaningfully describe user preferences, while assembling specific, granular domain ontologies to represent a general domain again elevates the concept space to an unmanageable size.

The lack of an expressive ontology that adequately describes the broad networked media domain from the users' perspective but at the same time is not too abstract or too specific, in order to scale well and maintain the decidability of reasoning algorithms has urged us to engineer LUMO (Section 2), and means to render it inter-usable with prominent LOD vocabularies (Section 3). Future work is presented in Section 4.

¹ www.linkedtv.eu

² <http://schema.org/>

2 LUMO: a lightweight reference ontology

LUMO's modelling objective is to represent user preferences and contextual features regarding networked media content in order to enable intelligent recommendations over concepts and content via semantic inferencing.

Design principles. LUMO aims to provide a comprehensive coverage of the domain through a uniform, finite and expressive vocabulary which considers the user's perspective of the domain, while at the same time remains as lightweight as possible so that it can enable user models' storage and the recommendation process to be performed on the user client, even if that consists of a limited-resource device. The latter aims to alleviate scalability issues but, most prominently, to safeguard user privacy by limiting the need to communicate sensitive user information to remote servers. The design methodology used bares resemblance to Methontology[2], with an emphasis on inspection of similar ontologies and vocabularies for knowledge acquisition.

Description of LUMO. LUMO³ is designed as an OWL 2 RL [3] ontology, combining ontologies and rules. This profile limitation offers a tradeoff between expressivity and reasoning efficiency in order to reduce computational cost. LUMO is currently accompanied by a separate ontology⁴, modelling mappings of LUMO to several existing vocabularies. Both⁵ were developed using Protégé⁶.

The current version of LUMO includes 804 classes and 6 object properties. It addresses four major user-pertinent facets: *Context Dimensions*, *Domain-dependent dimensions*, *Agents* and *Spatio-Temporal* aspects (Figure 1). The two latter facets may regard both contextual and domain-dependent aspects of the user and as so are modelled independently at the top of the hierarchy.

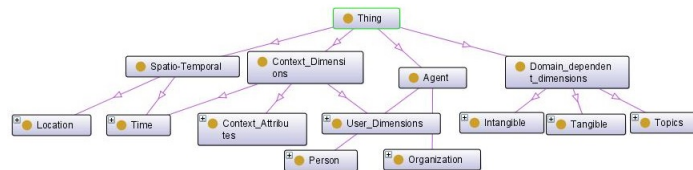


Fig. 1. Graphical representation of the top level LUMO hierarchy

Apart from the taxonomy, LUMO incorporates roughly 200 universal restriction axioms, relating concepts via non-taxonomical object properties. In the current version, such axioms are mostly of the type $Type \sqsubseteq \forall has(Sub)Topic.Topic$ and enable inference of topics in both the content's annotation and in the user profile, since content annotation in LinkedTV does not provide this information. *Type* corresponds to the type of an entity in the annotation (usually an agent, event, location or object), *Topic* is subsumed by the *Topics* concept/category and *hasSubTopic*, *hasTopic* are object properties. An example can be seen in

³ <http://data.linkedtv.eu/ontologies/lumo>

⁴ http://data.linkedtv.eu/ontologies/lumo_mappings

⁵ also available at <http://mklab.iti.gr/project/lumo>

⁶ <http://protege.stanford.edu/>

Figure 2.

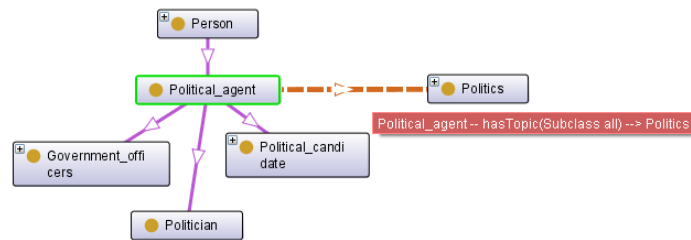


Fig. 2. An illustration of the relation of concept *Political_Agent* to the topic *Politics* via the *hasTopic* property

Influences from existing vocabularies. A key modelling objective was to adopt the most relevant entities and semantics from open vocabularies, discard user-extraneous information and adapt them to the requirements of the LinkedTV users and scenarios⁷. This included redefining semantics, with respect to leveraging LOD inconsistencies and enhancing coherency and completeness.

Within LinkedTV, multimedia content is annotated with entities from a plurality of LOD vocabularies describing *what a fragment is about*. Representing user-relevant information from consumed content heavily relied on the vocabularies used in the annotations. To this end, schema.org, the DBpedia ontology⁸ and the NERD ontology⁹ influenced the modelling of the *Agent*, *Location*, *Intangible* and *Tangible* subhierarchies, while better structuring, adding new information and selecting the parts that complied with the desired granularity was pursued.

The IPTC newscodes¹⁰ is widely used by news agents to categorize news content, thus relevant to the LinkedTV news scenario, therefore it consisted of the main influence in modelling the LUMO *Topics* subhierarchy. Most upper IPTC categories were adopted per se and subcategories and subsuming concepts were restructured and added.

The General User Model Ontology's (GUMO)[4] subsets that depict user context-relevant concepts such as user state and actions, e.g. motion, location, social environment, were adopted and adapted according to the specific LinkedTV requirements for context representation.

3 Mappings to existing vocabularies

LUMO's mappings serve a) for interpretation of content annotation to the LUMO vocabulary and b) as the means to facilitate re-use of LUMO by the Semantic Web. The mappings were generated automatically via the LogMap ontology alignment tool [6] and were further evaluated and revised manually.

⁷ News, artistic artifacts and a general artistic scenario: <http://www.linkedtv.eu/scenarios/>

⁸ <http://dbpedia.org/ontology/>

⁹ <http://nerd.eurecom.fr/ontology/>

¹⁰ <https://www.iptc.org/site/NewsCodes/>

Currently, mappings are available to the main vocabularies that influenced the engineering of LUMO: (a) the DBpedia ontology, (b) schema.org, (c) the NERD ontology, (d) IPTC news codes¹¹ and (e) the GUMO ontology¹². 524 LUMO classes are mapped to 785 classes from these vocabularies through equivalence axioms.

4 Conclusions and Future Work

This paper presented LUMO and its mappings to existing vocabularies. Future work will focus on evaluating and enriching LUMO's coverage over the generic networked media domain and also expand across the arts subdomain as a requisite of LinkedTV scenarios. We will also endeavour to engineer more non-taxonomical axioms (such as connection of objects/events to genres) and extend to more object properties. Lastly, extending the mappings to more prominent vocabularies is a major objective. Already under way are mappings to respective parts of YAGO and to a dedicated set of entities within DBpedia resources which are deemed semantically high-level enough to consist of concepts at the schema level and are not in the DBpedia ontology.

Acknowledgments

This work has been supported by the European Commission under Contract FP7-287911 LinkedTV.

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.* 5(3), 1–22 (2009)
2. Fernández-López, M., Gómez-Pérez, A., Juristo, N.: Methontology: from ontological art towards ontological engineering. *AAAI97* (1997)
3. Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P.F., Sattler, U.: OWL 2: The next step for OWL. *J. Web Sem.* 6(4), 309–322 (2008)
4. Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., von Wilamowitz-Moellendorff, M.: Gumo - the general user model ontology. In: *User Modeling*. pp. 428–432 (2005)
5. Jain, P., Hitzler, P., Yeh, P.Z., Verma, K., Sheth, A.P.: Linked data is merely more data. In: *AAAI Spring Symposium: Linked Data Meets Artificial Intelligence* (2010)
6. Jiménez-Ruiz, E., Grau, B.C.: Logmap: Logic-based and scalable ontology matching. In: *International Semantic Web Conference (ISWC)*. *Lecture Notes in Computer Science*, vol. 7031, pp. 273–288. Springer (October 2011)
7. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal* (2014)
8. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: A core of semantic knowledge. In: *Proceedings of the 16th International Conference on World Wide Web*. pp. 697–706. WWW '07, ACM, New York, NY, USA (2007)
9. Tsatsou, D., Mezaris, V., Kompatsiaris, I.: Semantic personalisation in networked media: Determining the background knowledge. In: *SMAP*. pp. 101–106 (2012)

¹¹ As owl-fied by the WebTlab (<http://webtlab.it.uc3m.es/>): <http://webtlab.it.uc3m.es/results/NEWS/subjectcodes.owl>

¹² <http://www.ubisworld.org/ubisworld/documents/gumo/2.0/gumo.owl>