OntoLife: an Ontology for Semantically Managing Personal Information

Eleni Kargioti, Efstratios Kontopoulos and Nick Bassiliades

Dept. of Informatics, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

{elkar, skontopo, nbassili}@csd.auth.gr

Abstract Personal knowledge management has been studied from various angles, one of which is the Semantic Web. Ontologies, the primary knowledge representation tool for the Semantic Web, can play a significant role in semantically managing personal knowledge. The scope of this paper focuses on addressing the issue of effective personal knowledge management, by proposing an ontology for modelling the domain of biographical events. The proposed ontology also undergoes a thorough evaluation, based on specific criteria presented in the literature.

1   Introduction

The latest technological developments and the WWW expose users to a great volume of information. A new perspective in *Knowledge Management* (*KM*) is essential that will filter out irrelevant information and increase knowledge quality, by utilizing the underlying semantic relationships. This requirement is also present in *Personal Knowledge Management* (*PKM*).

The first step towards PKM is to organize personal information. Various tools and applications are used (e.g. task managers, spreadsheet applications), but often comprise isolated solutions, revealing the need for a unified way of managing personal information, so that it becomes knowledge. *Ontologies* can assist towards this direction. They are a key factor towards realizing the *Semantic Web* vision [1], which promises to structure and semantically annotate raw information, to allow its interoperability, reuse and effective search by non-human agents.

This paper focuses on the issue of semantically managing the great volume of personal information by the use of an appropriately defined ontology. More specifically, an ontology called OntoLife is proposed for describing a person's biographical events and personal information. The ontology underwent a thorough evaluation that indicates its suitability for the designated purpose.

The rest of the paper is organized as follows: Section 2 describes related work paradigms, while the next section focuses on the presentation of the proposed on-

tology, accompanied by its evaluation. The paper concludes with final remarks and directions for future work.

2   Ontologies and the Semantic Web

Ontologies are the primary knowledge representation tool in the Semantic Web [1]. An ontology is a structured representational formalism of a domain, including a set of domain concepts and the relationships between them. The concepts describe classes of objects, while the relationships describe hierarchical dependencies among the concepts.

Regarding the domain of "life", the authors are not aware of an existing appropriate ontology. The *FOAF*[1] ontology is relevant, yet not wide enough for our purposes. *ResumeRDF*[2] is another ontology for representing Curriculum Vitae information about work and academic experience, skills, etc. Finally, another paradigm is *HR-XML*[3], a library of XML schemas that a variety of business processes related to human resource management. Nevertheless, none of the above (or other) ontologies and schemas can cover so broadly all the aspects of a person's biographical events as OntoLife.

3   Proposed Ontology

The scope of the proposed ontology is to model life by describing the person's characteristics, relationships and experiences. Since the domain is broad, an attempt to model it in details would produce a huge and cumbersome ontology. Thus, the domain is modelled in a non-exhaustive yet sufficient way, adopting the definition of generic entities that can easily be extended.

3.1 Description

The backbone of OntoLife is the *Person* entity. The entire ontology is built upon and around a Person, by a set of properties that relate the Person with the rest of the entities, as shown in Fig. 1. At the same time, many auxiliary entities and properties are defined to further describe the domain.

When designing the ontology, the idea of reusing commonly accepted ontologies was always considered. Thus, the Person entity of the FOAF ontology

---

[1] Friend Of A Friend (FOAF) ontology: `http://xmlns.com/foaf/0.1/`

[2] ResumeRDF Ontology: `http://rdfs.org/resume-rdf/`

[3] HR-XML: `http://ns.hr-xml.org/2_5/HR-XML-2_5/SEP/Candidate.xsd`

(foaf:Person) was enriched with new properties. Also, classes from external ontologies were imported, to model specific sub-domains in detail. These are the ISO lists for countries and languages[4] and the *Publication* and *Project* classes[5].
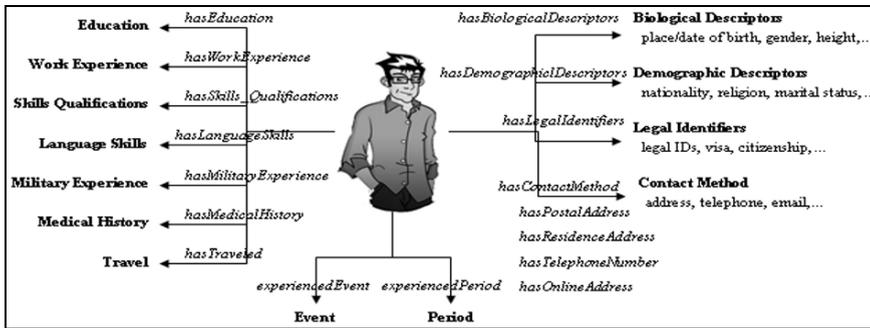

Fig. 1. Object properties with Person as the domain

OntoLife models the Person's demographics, biological and legal descriptors, the various contact methods and online accounts as well as information concerning educational, work and other experiences, qualifications and skills acquired and other CV-related information. Additionally, the link between a person and the periods and events experienced is thoroughly modelled.

Moreover, the Person entity is further extended by various subclasses, defined using external ontologies. Properties defined by the Relationship[6] ontology, which describes the relations between entities of foaf:Person are re-used. Thus, subclasses of Person, such as Friend, Colleague, Tutor, etc, are defined using restrictions on properties. Also, the external ontology Family Tree[7] that describes the domain of Family is imported. This ontology defines subclasses of foaf:Person, such as Child, Parent, Aunt/Uncle etc., using restrictions on related properties.

Two other important entities are *Period* and *Event*. Both are used to organise the biographical experiences in time periods and events and are linked with Person with properties inverse to the properties *experiencedPeriod* and *experiencedEvent* respectively. In further details, a *Period* models a time period and describes the place and dates it occurred, the people and organisations involved, related evaluations and also the events occurred within that period, using the inverse property *relatedEvent* (*duringPeriod*). Direct subclasses are Education, WorkExperience, Travel, MilitaryExperience and MedicalHistory. Similarly, an *Event* models a specific moment in life that can be part of a period or not. The place and date it occurred, the people and organisations involved, related evaluations, weather condi-

---

tions, scope and possible periodicity are described. Direct subclasses are PurchaseEvent, MedicalExaminationEvent and FamilyEvent. Further properties are defined with domain the subclasses of entities Period and Event and enable a more precise annotation of the related content. Fig. 2 describes in detail these subclasses and properties.
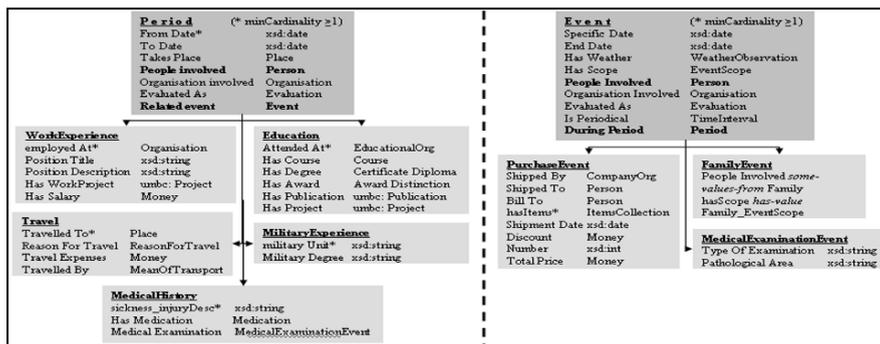

Fig. 2. Properties and subclasses of Period and Event

Finally, the ontology defines a super-class "Resource" which is the domain of four additional properties, *name*, *description*, *is Current*, *mime File* and which purpose is to pass these properties to all other classes, except Person and the Enumeration classes, where these properties are not needed. Resource is disjoint with Person and all auxiliary Enumeration classes.

3.2 Evaluation

The increasing number of ontologies in the web has led research to define methods and measures for evaluation. A popular method, also used in this work, is the *criteria-based evaluation* [3]. Consistency, completeness, expandability, minimal ontological commitment, etc. are some of the criteria listed in literature.

The adopted evaluation methodology includes the definition of specific requirements that the ontology needs to satisfy and the mapping of each requirement to a criterion [4]. Suitable measures are then selected to quantitatively assess each requirement. The main requirement is to be easy and intuitive to annotate content based on the ontology. Towards this, more specific requirements need to be met.

1. The terms used for the class names need to be close to real life terminology.
2. The classes should have a balanced number of subclasses; sufficient enough to facilitate effective annotation, but not too high to confuse the user.
3. The ontology should be rich, concerning attributes and relationships.
4. Cycles and other errors in the ontology structure should be avoided.

The Semantic Quality [5] criterion is mapped to the first requirement and the following measures. *Interpretability*, which is the percentage of class names that have a definition listed in *WordNet*[8] and *Concept Paths* [6], which is the percentage of class hierarchies that are depicted in WordNet through term hyponyms. Expandability/Coverage [7] are the criteria mapped to the second requirement. Related measures are: class tree depth, breath and branching factor. For a broad ontology like OntoLife, a less deep tree with a low branching factor is preferred. The third requirement is mapped to the ontology richness criterion [8], assessed by the attribute and relationship richness. The last requirement is mapped to the Minimal Ontological commitment criterion [9] and ontology validators are used to exclude circularity and other types of errors.

To assess semantic quality, certain assumptions are made. Class names that consist of more than one word written in CamelCase or separated by underscore were considered listed in WordNet, if all included words were listed (e.g. Certificate_Diploma) or a phrase with these words made sense (e.g. ForeignLanguage). Also, class names taken from the HR-XML Candidate specification were considered interpretable. Finally, a concept path (the path from a parent class to a leaf subclass in a class tree) may be fully or partially depicted in WordNet, if all or some subclasses are listed as parent class hyponyms. To measure Interpretability and Concept Paths, a weighted average was calculated (Table 1). To assess ontology expandability/coverage and attribute and relationship richness, the metrics offered by the SWOOP[9] ontology editor were used (Table 2). Finally, to assess the last criterion the *Vowlidator*[10] and *WonderWeb online validator*[11] were used. The ontology was identified as OWL Full compatible, while no errors were indicated.

Table 1. Calculation of measures for the Semantic Quality criterion

Interpretability: (c1 x 0 + c2 x 0.5 + c3 x 1) / c = 66%
c1: no of classes whose sense is not listed in WordNet,
c2: no of classes with a sister/ synonym term
c3: no of classes with a sense listed in WordNet,
c: total no of defined classes
Concept Paths: (p1 x 0 + p2 x 0.5 + p3 x 1) / p = 60%
p1: no of concept paths not depicted by WordNet
p2: no of concept paths partially depicted by WordNet
p3: no of concept paths depicted by WordNet,
p: total no of concept paths

---

[8] WordNet, a lexical database by the Princeton University Cognitive Science Laboratory

[9] SWOOP: A Hypermedia-based Featherweight OWL Ontology Editor

[10] Vowlidator: `http://projects.semwebcentral.org/projects/vowlidator/`

[11] WonderWeb OWL Validator: `http://www.mygrid.org.uk/OWL/Validator`

　The measures show room for improvement in the class name terminology. Nevertheless, 66% and 60% for interpretability and concept paths are satisfactory, especially for such a broad ontology. Also, the number of classes is right based on the measures of the expandability/coverage. For a tree of such breadth, the depth and branching factor are low, resulting in a rather simple class hierarchy. Moreover, the ontology is rich enough concerning attributes and relationships permitting detailed content annotation. Finally, the ontology can be used "safely", since syntactical errors and inconsistencies were excluded.

Table 2. Calculation of measures for the ontology's expandability/coverage criteria and the attributes and relationship richness criterion

Tree Depth
Max. Depth of Class Tree: 4, Min. Depth of Class Tree: 1, Avg. Depth of Class Tree: 1.9
Tree Breadth
Max. Breadth of Class Tree: 33, Min. Breath of Class Tree: 1, Avg. breadth of Class Tree: 25
Tree Branching factor
Max. Branching Factor of Class Tree: 47, Min. Branching Factor of Class Tree: 1 *Avg. Branching Factor of Class Tree: 6.6*
Attribute richness
No. Attributes in all classes / No. classes = 85%
Relationships richness
No. Relations / (No. Subclasses+No. Relations) = 68%

4　Conclusions and Future Work

The paper argued that Personal Knowledge Management is increasingly gaining attention, to facilitate the end-user in handling vast volumes of information. Ontologies and the Semantic Web can support this task, by offering capabilities for semantically managing personal information via properly defined ontologies. Such an ontology is proposed in this work. The ontology, called OntoLife, portrays a person's life, by describing his/her characteristics, his/her relationships with other people and the various events experienced. Since the domain is so broad, the ontology is modelled in a non-exhaustive way, defining generic entities that can easily be extended. OntoLife also underwent a thorough evaluation that indicated its suitability for the designated purpose.

　Future research may include a revision of the ontology, which could incorporate the new version of FOAF or of other external ontologies that model a sub-domain in detail. Furthermore, it would be interesting to study the combination of the proposed ontology with an appropriate tool for managing ontologies, such as a general-purpose Semantic Wiki. Semantic Wikis enrich standard Wikis with functionalities deriving from the content's semantics and aim at knowledge reuse, adaptive interface and navigation and effective search [10]. The purpose of the study

would be on one hand to evaluate how efficiently Semantic Wikis can incorporate and represent the proposed ontology and on the other hand to see how effectively they can support users in order to annotate content, even when with minimal knowledge of the underlying ontology.

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American, 284(5):34-43, 2001.
2. Staab, S., Studer, R.: Handbook on Ontologies. International Handbooks on Information Systems, Springer Verlag (2004)
3. Gomez-Perez, A.: Evaluation of Ontologies. International Journal of Intelligent Systems, 16(3):391-409, 2001.
4. Yu, J., Thom, A.A., Tam, A.: Ontology Evaluation Using Wikipedia Categories for Browsing. In Proceedings 16th Conference on Information and Knowledge Management (CIKM), pp. 223-232 (2007)
5. Burton-Jones, et al: A Semiotic Metrics Suite for Assessing the Quality of Ontologies. Data Knowledge Engineering, 55(1):84-102, 2005.
6. Sleeman, D., Reul, Q. H.: CleanONTO: Evaluating Taxonomic Relationships in Ontologies. In Proceedings 4th International Workshop on Evaluation of Ontologies for the Web (EON), Edinburgh, Scotland (2006)
7. Gangemi, A. et al: A Theoretical Framework for Ontology Evaluation and Validation. In Proceedings 2nd Italian Semantic Web Workshop (SWAP), Trento, Italy (2005)
8. Tartir, S. et al: OntoQA: Metric-Based Ontology Quality Analysis. IEEE ICDM Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, Houston, TX (2005)
9. Yu, J., Thom, A.A., Tam, A.: Evaluating Ontology Criteria for Requirements in a Geographic Travel Domain. In Proceedings International Conference on Ontologies, Databases and Applications of Semantics (2005)
10. Leuf, B., Cunningham, W.: The Wiki Way: Collaboration and Sharing on the Internet. Addison Wesley, Reading, Massachusetts (2001)