

When Mommy Blogs are Semantically Tagged

Jinan El-Hachem and Volker Haarslev

Engineering and Computer Science Department
Concordia University, Montreal, Quebec, Canada
ji_elhac@encs.concordia.ca

Abstract. OWL 2-supported Semantic Tagging is a non compulsory yet decisive and highly influential component of a multidisciplinary knowledge architecture framework which synergetically combines the Semantic and the Social Webs. The facility consists of a semantic tagging layer based on OWL 2 axioms and expressions enticing social network users, typically mommy bloggers, to annotate their chaos of textual data with natural language verbalized versions of ontological elements. This paper provides a comprehensive short summary of the overall framework along with its backbone metamodel and its parenting analysis and surveillance ontology ParOnt, laying a particular emphasis on its semantic expression-based tagging feature, and accordingly highlighting the attained gains and improvements in terms of effective results, services and recommendations, all falling in the scope of public parenting orientation and awareness.

Keywords: Semantic Web, Social Web, Social Network Sites (SNS), Mommy Blog, Natural Language Processing (NLP), Natural Language Generation (NLG), Web Ontology Language (OWL), Description Logics (DL), Recommender Systems, Parenting, Public Awareness

1 Introduction

The emergence of tagging with Web 2.0 sites, and hence the increasing willingness of Social Network users to provide collaborative tagging make the availability of metadata with significant potentials an undisputable reality. Furthermore, these realized facts lead to the assessment that it would be simply unjustifiable if the opportunity of properly taking advantage of the potentials in question is missed through not fostering appropriate approaches to create and manage tags. Among the variety of accessible Web 2.0 sites, “Mommy blogs” are those in which Social Network users provide tons of information related to children, their problems and behaviors, to parenting in general; and “Mommy bloggers”, usually parents, are perceived as extremely active and cooperative users who constantly access and manage their blogs.

In this work, traditional semantic tagging notions are further developed to include identified rules and well defined OWL 2 ontology axioms. The paper thus briefly introduces a social semantic platform already proposed in the related previous works [5] and [6], and extends its semantic and rules tagging layer

using ontological formalisms and Natural Language Processing and Generation techniques to bestow effective and efficient ontology population, reasoning and querying facilities. The framework thus relies on highly expressive domain expert ontologies, namely ParOnt, a conceived ontology for parenting cross-sectional analysis. Its backbone metamodel fosters the different methods for semantic web language components' distinctive characteristics support, for instance OWL 2 profiles and sublanguages, providing projection mechanisms for efficient and substantiable performance support.

At the end, it demonstrates that if mommy bloggers are further involved in an efficient process of knowledge tagging, massive and efficient ontology population is easily achieved. Thus, supported by advanced reasoning and querying techniques, different services are offered, including community profiling and segmentation, generic and customized parenting recommendations, as well as alerts on encountered fallacies and misleading notions or beliefs, all in the scope of public parenting orientation and awareness.

The remaining part of this paper proceeds by presenting an overview of the knowledge framework and platform with the details of the incorporated semantic tagging feature. Following that, demonstrating scenarios and experimental examples that endow with recommender systems based on the Parenting Ontology ParOnt are exposed right before the concluding section that comprises a closing discussion along with highlights on some relevant future work.

2 Architectural Framework Overview: Emphasis on Semantic Tagging

Figure 1 below tries to overcome the page number limitations by presenting a simplistic high-level description of the knowledge architectural framework's main components. In this illustration, social mommy blog data are subject to an ontology aware analysis process, with the call of NLP and data mining techniques. Following that, a process of semantic annotations generation takes place, within a multidimensional approach that relies on prerequisite data reduction methods and ontology restrictions based on axioms and expressions' complexity levels. As a result, post-tagging ontology population is achieved. Tag suggestions facilities occur based on ontology experts concepts' annotations and axioms' natural language generation, leading to possible user certification and validation.

The populated ontology is subsequently accessible for the different reasoning engine techniques, including classification and subsumption, satisfiability and instance checking, inference discovery and query answering, rule validation and processing. These techniques are henceforth the means by which decision support systems capabilities are attainable.

The different parts of the process rely on the backbone repository. This repository encloses the domain ontologies as well as the metamodel encompassing

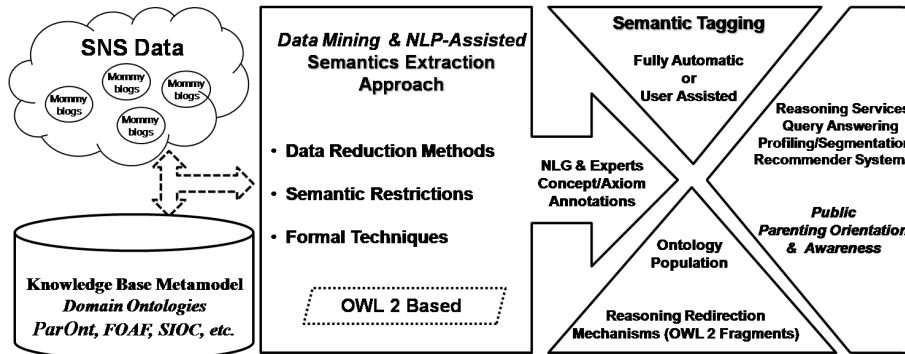


Fig. 1. Knowledge-based Architecture and Modeling Platform General Overview

“meta-semantics” structures to allow the recognition of sublanguages along with suitable reasoning services. Meta-semantics also play a role in the verbalization process of the different constructs prior to the user tagging phase, which additional details are explored next.

The primary formal definition of a tag was first provided in [7]. The definition was based on a tripartite model relying on an actor (a user), a concept (a tag or keyword), and an instance (annotated resource):

$$T \subseteq A \times C \times I \quad (1)$$

where A is for Actor, C for Concept and I for Instance.

Later on in [1], the tripartite model definition was extended to a quadripartite one, after adding a local semantic meaning to each tag, obtained by a URI:

$$Tagging \subseteq User \times Resource \times Tag \times Meaning \quad (2)$$

In this work, we further extend the above definition, by assigning a more granular element to the definition, using Description Logics (DL) to denote constructs, axioms and expressions.

Our definition can thus be denoted by the following:

$$Tagging \subseteq User \times Resource \times Tag \times DL_{SEMANTICS} \quad (3)$$

where $DL_{SEMANTICS}$ are OWL 2 Constructs, Axioms and Expressions, in other words DL Building blocks forming OWL 2 fragments and languages, such as ALC and $SHOIN(D)$ [4, 3].

The framework thus promotes a syntactic and more formalized approach, benefiting from its metamodel repository’s already briefly introduced meta-semantics

structures that allow the distinguishing of constructs such as existential restrictions, class conjunctions, disjunctions and negations, cardinality restrictions, ranges and datatypes, nominals, role properties (inverse, transitive, hierarchical, and so on).

The motivation behind attaining this level of granularity is to overcome data mining and NLP limitations by reusing ontology definitions and rules tagging, enforced by a possible user cooperation, thus cutting down complex algorithms and compensating for them through reasoning.

Being aware of the fact that non-ontology experts will surely face difficulties trying to read and understand ontology elements, natural language definitions of classes and axioms are made available based on ontology and metamodel prepared annotations on one hand, and on NLG-based techniques (such as On-to-Verbal ¹) on the other. The SN user is thus faced with verbalised naturalistic versions of the formal semantics.

3 Proof of Concept, Experimental Application Scenarios

The first step leading to the platform’s realization is the design of its metamodel, with the meta-semantics structures, along with all inferred patterns and configurations. The main ontology conceived for the purpose of parents’ orientation and awareness is OWL 2’s ParOnt (Parenting Ontology), an upgraded integrated version of COPE(Childhood Obesity Prevention [Knowledge] Enterprise) [2, 6], with source information derived from trustful data sources such as RAMQ ², Canadian Community Health Survey (CCHS)³ (population health database), CARTaGENE ⁴.

ParOnt’s major subdomains include baby development, behavior, nutrition, health, safety, activities, etc.

Useful Social Network Sites (SNS) data sources typically beneficial for our parenting domain are “Mommy Bloggers”, such as Babycenter ⁵ (which alone counts more than 20 million users), Canada Moms Blog⁶, Raising Children Network⁷, Asian Mommy ⁸, among others. The study reported in this paper is based on 2000 blogs and replies collected from these SNS, falling within the same period interval (between 2011 and 2012). A group of almost 60 taggers was in charge of regularly annotating textual blogs with DL semantics, submitted to them in the form of short true false questionnaires.

While the full strategy to process blog data (based on semantics for domain

¹ <http://swatproject.org/demos.asp>

² Régie de l’assurance maladie du Québec: www.ramq.gouv.qc.ca/index_en.shtml

³ www.statcan.gc.ca/concepts/health-sante/index-eng.htm

⁴ www.cartagene.qc.ca/index.php?lang=english

⁵ www.babycenter.com

⁶ www.canadamommsblog.com

⁷ www.raisingchildren.net.au

⁸ www.asianmommy.com

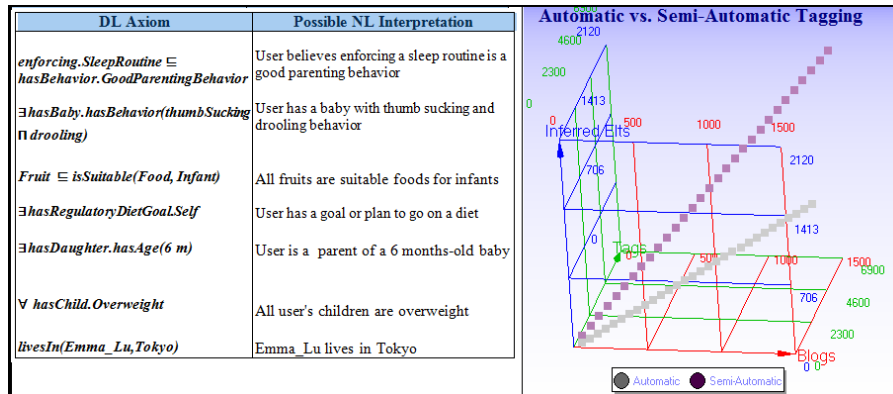


Fig. 2. To the left, examples of DL constructs, along with their possible Natural Language interpretation; to the right, a comparative graph highlighting the difference between automatic and semi-automatic approaches for tagging/ ontology population. Note that the more we add complexity, the more NLP and mining techniques become limited, and the more the user tagging role is highlighted and rendered crucial.

and rule dedicated search on one hand, and filtering criteria for data reduction on the other [5]) are not the focus of this paper, a flat straightforward workflow was adopted to provide a minimum level of support to show the preliminary advantage of a semi-automatic ontology population process.

Figure 2 provides sample ontology axioms along with their verbalized definitions, as well as a comparative graphical illustration denoting, based on OWL DL ontology subsets and sample filtered blog extracts, the advantage of a semi-automatic user assisted population process. Users are offered a facility through which the enrichment of the ontology with instances goes beyond automatic NLP and data mining capabilities. This naturally increases the number of inferred elements, which represent query answering, inference discovery, instance checking and association to additional ontology classes (taking into consideration the multidimensional nature of an ontology). It is worth noting that the estimations provided (for inferred elements) are not static or fixed numbers. The idea is to prove that for a given number of analyzed blogs, the number of tags easily doubles between an automatic and a semi-automatic approach, proportionally with the number of inferred new knowledge, of services, etc. Furthermore, the more the level of expressivity and complexity increases, the steeper the slope will be, as extra automatic tagging restrictions are reinforced: with more complex semantics, accuracy levels are aggravated; as a consequence, the positive input brought in by the user's tagging cooperation is exponentially valued.

Once this infrastructure is achieved, reasoning procedures can be applied in order to attain the required services for our parenting awareness and orientation systems. Redirection mechanisms, based on the projected languages and frag-

ments, target advanced and powerful reasoners, query and rule engines. Useful retrieved information, mostly based on conjunctive queries, bring in recommender services, profiling, segmentation, opinion mining capabilities; more concrete examples include: the retrieval of the list of frequent undesired behaviors, of health problems for children per age in certain locations, of sets of best practice recommendations for parenting in general or in specific situations, of detected bad parents' behaviors, misconceptions, and so on.

4 Conclusions

The main contributions of this paper consist in the following:

- It provided highlights on an already proposed framework surrounding its own main contribution
- It put forward an extension of the tagging semantic model quadripartite model
- It introduced ParOnt, a newly conceived Parenting Ontology
- It demonstrated the overall advantages and services of getting SN users more involved in a formalized tagging process, all falling under the scope of public parenting orientation and awareness

In terms of future work, we plan to further develop implementation and verification tools, looking for the incorporation of maximized sets of rules and Description Logics-based languages.

References

1. A. Passant, P. Laublet. Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In Proceedings of the WWW2008 Workshop Linked Data on the Web (LDOW2008), Apr 2008
2. A. Shaban-Nejad, D. Buckeridge, D. L. Dub, COPE: Childhood Obesity Prevention [Knowledge] Enterprise. in Proceedings of AIME 2011, pp. 225-229
3. B. Cuenca Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider and U. Sattler, "OWL 2: The next step for OWL", *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, no. 4, pp. 309-322, 2008
4. F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider, editors, "The Description Logic Handbook: Theory, Implementations and Applications", *Cambridge University Press*, 2007.
5. J. El-Hachem, V. Haarslev. A User and NLP-Assisted Strategic Workflow for a Social Semantic OWL 2-Based Knowledge Platform. In: Semantic Analysis in Social Media, at EACL 2012, April 2012
6. J. El-Hachem, A. Shaban-Nejad, V. Haarslev, L. Dube, D. Buckeridge. An OWL 2-Based Knowledge Platform Combining the Social and Semantic Webs for an Ambient Childhood Obesity Prevention System. In: The International Conference on Ambient Systems, Networks and Technologies, *Elsevier*, August 2012
7. P. Mika, "Ontologies are us: A unified model of social networks and semantics", *Journal of Web Semantics, Elsevier*, Volume 5, Number 1, pp.5-15, 2007