

Challenges for Semantically Driven Collaborative Spaces

Pascal Molli¹, John G. Breslin², and Maria-Esther Vidal³✉

¹ University of Nantes, Nantes, France
pascal.molli@univ-nantes.fr

² Insight Centre for Data Analytics, National University of Ireland Galway,
Galway, Ireland
john.breslin@nuigalway.ie

³ Universidad Simón Bolívar, Caracas, Venezuela
mvidal@ldc.usb.ve

Abstract. Linked Data initiatives have fostered the publication of more than one thousand of datasets in the Linking Open Data (LOD) cloud from a large variety of domains, e.g., Life Sciences, Media, and Government. Albeit large in volume, Linked Data is essentially read-only and most collaborative tasks of cleaning, enriching, and reasoning are not dynamically available. Collaboration between data producers and consumers is essential for overcoming these limitations, and for fostering the evolution of the LOD cloud into a more participative and collaborative data space. In this paper, we describe the role that collaborative infrastructures can play in creating and maintaining Linked Data, and the benefits of exploiting knowledge represented in ontologies as well as the main features of Semantic Web technologies to effectively assess the LOD cloud's evolution. First, the advantages of using ontologies for modelling collaborative spaces are discussed, as well as formalisms for assessing semantic collaboration by sharing annotations from terms in domain ontologies. Then, Semantic MediaWiki communities are described, and illustrated with three applications in the domains of formal mathematics, ontology engineering, and pedagogical content management. Next, the problem of exploiting semantics in collaborative spaces is tackled, and three different approaches are described. Finally, we conclude with an outlook to future directions and problems that remain open in the area of semantically-driven collaborative spaces.

1 Introduction

Over the last decade, there has been a rapid increase in the numbers of users signing up to be part of Web-based social networks. Hundreds of millions of new members are now joining the major services each year. A large amount of content is being shared on these networks, and around tens of billions of content items are shared each month. In parallel, similar collaborative spaces are being leveraged in both private intranets and enterprise environments; these collaborative spaces have features mirroring those on the public Web.

With this growth in usage and data being generated, there are many opportunities to discover the knowledge that is often inherent but somewhat hidden in these networks. Web mining techniques are being used to derive this hidden knowledge. In addition, Semantic Web technologies, including Linked Data initiatives to connect previously disconnected datasets, are making it possible to connect data from across various social spaces through common representations and agreed upon terms for people, content items, etc.

In this volume, we will outline some current research being carried out to semantically represent the implicit and explicit structures on the Social Web, along with the techniques being used to elicit relevant knowledge from these structures, and the mechanisms that can be used to intelligently mesh these semantic representations with intelligent knowledge discovery processes.

2 Modelling Collaborative Communities and the Role of Semantics

Semantics represented in ontologies or vocabularies can be used to enhance the description and modelling of collaborative spaces. In particular, annotating data with terms from ontologies is a common activity that has gained attention with the development of Semantic Web technologies. Scientific communities from natural sciences such as Life Sciences have actively used ontologies to describe the semantics of scientific concepts. The Gene and Human Phenotype Ontologies have been extensively applied for describing genes, and there are international initiatives to collaboratively annotate organisms, e.g., the *Pseudomonas aeruginosa* PAO1 genome¹. One of the main goals to be achieved to support a precise modelling in collaborative spaces is the development of tools that conduct reasoning on top of existing ontologies and allow for the collaborative annotation of these entities.

In this direction, Goy et al. [3] propose a model to represent views of a data set and different versions of annotations of the data enclosed in the view. Annotations can be personal, allowing for the representation of individual conceptualisations of the portion of the domain represented by the view. Additionally, annotations can be shared in case the annotations can be visible to all members of the collaborative space. The proposed model is part of the project *Semantic Table Plus Plus* (SemT++) [4] which provides a platform to collaboratively describe web resources, i.e., to semantically describe images, documents, videos, or any other resource publicly available on the Web. Existing ontologies are provided to model knowledge about resources represented using the ontology annotations, e.g., the DOLCE² and Geographic ontologies³ provide controlled vocabularies to describe Web resources. The benefits of using personal annotations are illustrated in a use case where documents are collaboratively described following an *authored collaboration policy*. This policy allows an authorised user

¹ <http://www.pseudomonas.com/goannotationproject2014.jsp>.

² <http://www.loa.istc.cnr.it/old/DOLCE.html>.

³ <https://www.w3.org/2005/Incubator/geo/XGR-geo-ont-20071023/>.

to delete other user annotations if there is no agreement across all users, while the annotations may remain in the user's local view.

3 Semantic MediaWiki Communities

For over ten years on Wikipedia, templates have been used to provide a consistent look to the structured content placed within article texts (these are called infoboxes on Wikipedia). They can also be used to provide a structure for entering data, so that it is possible to easily extract metadata about some aspect of an entity that is the focus of an article (e.g., from a template field called 'population' in an article about Galway). Semantic wikis bring this to the next level by allowing users to create semantic annotations anywhere within a wiki article's text for the purposes of structured access and finer-grained searches, inline querying, and external information reuse. These are very useful in enterprise scenarios when wikis are used as collaborative spaces where structured data can be easily entered and updated by a distributed community of enterprise users. One of the largest semantic wikis is Semantic MediaWiki, based on the popular MediaWiki system.

In the context of domain-specific applications, Kaliszyk and Urban [5] provide an overview of collaborative systems for collecting and sharing mathematical knowledge. One of the problems to be solved by these systems is the visualisation of formal proofs for people, while providing assistance with the translation of informal statements into formal mathematics. Different problems regarding disagreements between mathematicians, e.g., in terms of the lack of a unique formal language and axiomatic systems, have complicated the development of such formalisation frameworks, necessitating collaborative work like the one implemented in wiki applications such as Wikipedia. These collaborative systems or formal wikis allow for formal verification, the implementation of formal libraries to provide a unified terminology and theorems, versioning of the proofs, semantic resolution of ambiguities, collaborative editing, and other semantic tools. The authors illustrate some of these features in their systems. The Mizar Wiki [1] provides a mathematical library that includes theorems that can be reused in other proofs; additionally, Mizar makes available a proof checker to validate the proposed proofs and to facilitate a peer review process. Furthermore, ProofWeb is a web interface for editing pages and assisting users during theorem demonstration. The authors conclude that there are many challenges to be achieved before collaborative systems in formal mathematics can become a reality. In particular, semantics can play a relevant role in future developments.

A paper by Rutledge et al. [7] introduces a technique to annotate and browse a given ontology based on *Fresnel forms*⁴. Fresnel forms are created via a Protegé plugin that allows users to edit and reuse the ontologies, and also to specify how the semantics should be browsed/displayed (using CSS to create a suitable visualisation for a given data structure). The result is a useful tool that exploits the semantics encoded in an ontology to generate semantic wikis. In particular,

⁴ http://is.cs.ou.nl/OWF/index.php5/Fresnel_Forms.

the authors apply this technique to bridge Semantic MediaWiki with a browsing interface and semantic forms. Evaluation of this approach was carried out in a case study where the authors discussed the feasibility of implementing their tool via its application to the well-known FOAF ontology, and its usefulness in a Wikipedia infobox-style interface.

Zander et al. [9] tackle the problem of pedagogical content management and evaluate how semantic collaborative infrastructures like semantic wikis can have a positive impact on content reusability and authoring. The novelty of this approach relies on an extension to semantic wikis with knowledge encoded in ontologies like the Pedagogical Ontology (PO) and the Semantic Learning Object Model (SLOM⁵) to enhance expressiveness and interoperability across multiple curricula and pedagogies. Furthermore, the combination of these technologies facilitates the generation of pedagogical knowledge as Linked Data. The effectiveness of the proposed framework is empirically evaluated with a user study and measured using a usability test. Reported results suggest that non-computer specialists are highly satisfied with semantic wikis enhanced with pedagogical knowledge, facilitating the task of creating semantically enriched content for teaching and learning.

4 Exploiting Semantics in Collaborative Spaces

Semantics encoded in ontologies and Linked Data sets can have a positive impact on the behaviour of applications built on top of collaborative spaces. The main challenge to be addressed is how to efficiently extract the knowledge represented in existing Linked Data sets and effectively use this knowledge to enhance existing collaborative spaces. Improvements can be of diverse types, e.g., inferred facts from DBpedia can be used for quality assessment of collaborative space contents; thus, diverse methods to uncover data quality problems should be defined. Additionally, queries against these Linked Data sets have to be crafted in such a way that query answers will provide the insights to discover the missing or faulty content in the collaborative space. Finally, evaluation methodologies are required to determine the quality of the discoveries and to precisely propose changes that will assess high quality content of collaborative spaces.

This special issue compiles two exemplar applications where the benefits of using semantics are clearly demonstrated. First, Torres et al. [8] present BlueFinder, a recommendation system able to enhance Wikipedia content with information retrieved from DBpedia⁶. BlueFinder identifies the classes to which a pair of DBpedia resources belong to, and uses this information to feed an unsupervised learning algorithm that recommends new associations between disconnected Wikipedia articles. The behaviour of BlueFinder is empirically studied in terms of the number of missing Wikipedia connections that BlueFinder can detect; reported results suggest that by exploiting the semantics encoded in DBpedia, BlueFinder is able to identify 270,367 new Wikipedia connections.

⁵ <http://www.intuitel.de/>.

⁶ <http://wiki.dbpedia.org/>.

BlueFinder addresses the challenges of uncovering missing content in Wikipedia and designing DBpedia queries to recover missing values. As a future direction, the authors propose extending BlueFinder to allow collaborative validation of these values via crowdsourcing and to include the discovered links in Wikipedia.

Following this line of research, Louati et al. [6] tackle the problem of aggregating heterogeneous social networks, and propose a hybrid graph summarisation approach. The proposed technique extends existing clustering methods such as K-medoids and hierarchical clustering to cluster heterogeneous social networks. The novelty of this approach relies on the usage of attributes and relationships represented in the network, to produce clusters that better fill user requirements. The proposed summarisation techniques implement an unsupervised learning algorithm that uses Rough Set Theory to enhance the precision of known clustering methods: K-medoids and hierarchical. The proposed approach is empirically evaluated in an existing social network data set on US political books⁷. The results suggest that exploiting the semantics encoded in the attributes and relationships positively impacts on the quality of the communities identified by the proposed techniques. The challenges achieved by the authors provide the basis for the development of tools for uncovering patterns of the data that suggest quality problems in the content of the collaborative space. In the future, the authors plan to extend the proposed graph summarisation techniques to dynamically adapt the attributes and relationships used to cluster the input network. This extension will allow for the representation of more general domain restrictions, and in consequence, a more expressive semantically-driven clustering approach.

If Semantic Web technologies have allowed many data providers to publish datasets with their own ontologies, consuming these datasets requires aligning different ontologies and performing entity matching. The Semantic Web community has developed sophisticated tools to tackle this problem, but a generic automatic reliable solution is still an open research problem. Human-machine collaboration can be effective to promote reuse of contextual trusted ontology mappings. This special issue also includes work by Bottali et al. [2] on Okkam, a collaborative tool designed to share ontology mappings. Okkam allows users to disagree on mappings, but also to rank mappings according to their *contextual sharedness*. Consequently, Okkam allows users to quickly find mappings adapted to their usage, knowing their level of agreement for a given particular context.

5 Future Directions

A large research effort has enabled the continuous transformation of content to knowledge, building the Semantic Web from a Web of Documents and the Deep Web. A new major research challenge is to ensure a co-evolution of content and knowledge, making both of them trustable. This means that we must be able to not only extract and manage knowledge from contents, but also to augment contents based on knowledge. Co-evolution of content and knowledge happens in a semantic wiki, but the question is how can we scale this at the levels of the Web of Documents and the Web of Data.

⁷ <http://www-personal.umich.edu/~mejn/netdata/>.

Co-evolution of content and knowledge is a powerful mechanism to improve both content and knowledge, but such an evolution has to be powered by a new form of collaboration between humans and AI. From the point of view of people, human-AI collaboration means that we must make formal knowledge and its evolution accessible, usable, editable and understandable, so that people can observe, control, evaluate and reuse this formal knowledge. All research works related to explanations and knowledge revision follow this direction.

In the other direction, from the point of view of computers, human-AI collaboration means that we must be able to take into account the unpredictable behaviours of people who can at any moment in time add or modify content and formal knowledge, with the risk of introducing uncertainty or inconsistency. Research works related to uncertainty management in crowdsourcing and truth-finding algorithms are of interest here.

Bootstrapping the co-evolution of content and knowledge requires that we find new ways for humans and AI to collaborate. Knowledge benefits, i.e. querying, reasoning, fact checking, and truth finding, have to be easily available for any people using the Web. On the other hand, knowledge quality brings with it really challenging questions when deployed, for example, if the mistakes found by people contradict with (and should override) what is formally stated in accepted knowledge bases. Finally, the co-evolution of content and knowledge has to be monitored, ensuring that the overall quality of content and knowledge improves.

Acknowledgements. Pascal Molli has been supported by the ANR Kolflow Project (ANR-10-CORD-0021), University of Nantes. John G. Breslin has been supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (Insight). Maria-Esther Vidal thanks the DID-USB (Decanato de Investigación y Desarrollo de la Universidad Simón Bolívar).

References

1. Alama, J., Brink, K., Mamane, L., Urban, J.: Large formal wikis: issues and solutions. In: Davenport, J.H., Farmer, W.M., Urban, J., Rabe, F. (eds.) MKM 2011 and Calculemus 2011. LNCS, vol. 6824, pp. 133–148. Springer, Heidelberg (2011)
2. Bortoli, S., Bouquet, P., Bazzanella, B.: Okkam synopsis: connecting vocabularies across systems and users. In: Advances in Semantic Web Collaborative Spaces, Revised Selected Papers of the Semantic Web Collaborative Spaces (SWCS) 2013 and 2014 (2016)
3. Goy, A., Magro, D., Petrone, G., Picardi, C., Segnan, M.: Shared and personal views on collaborative semantic tables. In: Advances in Semantic Web Collaborative Spaces, Revised Selected Papers of the Semantic Web Collaborative Spaces (SWCS) 2013 and 2014 (2016)
4. Goy, A., Magro, D., Petrone, G., Segnan, M.: Collaborative semantic tables. In: Proceedings of the Third International Workshop on Semantic Web Collaborative Spaces Co-located with the 13th International Semantic Web Conference (ISWC), Riva del Garda, Italy, 19 October 2014

5. Kaliszzyk, C., Urban, J.: Wikis and collaborative systems for large formal mathematics. In: *Advances in Semantic Web Collaborative Spaces, Revised Selected Papers of the Semantic Web Collaborative Spaces (SWCS) 2013 and 2014* (2016)
6. Louati, A., Aufaure, M.-A., Cuvelier, E., Pimentel, B.: Soft and adaptive aggregation of heterogeneous graphs with heterogeneous attributes. In: *Advances in Semantic Web Collaborative Spaces, Revised Selected Papers of the Semantic Web Collaborative Spaces (SWCS) 2013 and 2014* (2016)
7. Rutledge, L., Brenninkmeijer, T., Zwanenberg, T., van de Heijning, J., Mekkering, A., Theunissen, J.N., Bos, R.: From ontology to semantic wiki – designing annotation and browse interfaces for given ontologies. In: *Advances in Semantic Web Collaborative Spaces, Revised Selected Papers of the Semantic Web Collaborative Spaces (SWCS) 2013 and 2014* (2016)
8. Torres, D., Skaf-Molli, H., Molli, P., Diaz, A.: Discovering wikipedia conventions using DBpedia properties. In: *Advances in Semantic Web Collaborative Spaces, Revised Selected Papers of the Semantic Web Collaborative Spaces (SWCS) 2013 and 2014* (2016)
9. Zander, S., Swertz, C., Verdu, E., Perez, M.J.V., Henning, P.: A semantic mediawiki-based approach for the collaborative development of pedagogically meaningful learning content annotations. In: *Advances in Semantic Web Collaborative Spaces, Revised Selected Papers of the Semantic Web Collaborative Spaces (SWCS) 2013 and 2014* (2016)