The 2nd U.S. Semantic Technologies Symposium (US2TS 2019)

Editorial

Anna Lisa Gentile [a], Tania Tudorache [b]
[a] *IBM Research, CA, USA*
E-mail: annalisa.gentile@ibm.com
[b] *Stanford University, CA, USA*
E-mail: tudorache@stanford.edu

Abstract. This report provides an overview of the 2nd edition of the U.S. Semantic Technologies Symposium (US2TS) series that took place between March 11-13, 2019 at Duke University, Durham, North Carolina. The main goal of the US2TS symposium series is to facilitate the formation of a coherent national agenda for exploring emerging trends in Semantic Technologies, and to help consolidate and build a U.S.-based community research network. This report describes the structure, program and the outcomes of the second edition of US2TS (http://us2ts.org/2019/).

1. The US2TS Series

The Semantic Web is an inherently multi-disciplinary field. The Artificial Intelligence community has contributed much in the way of formal logic and knowledge representation. Similarly, the applied computer science community, along with industry and government agencies, have contributed with application development and testing. With an ever-growing dependence on the web, and the continuously increasing importance of large-scale data sharing, integration, and reuse, natural science researchers, geoscience, biology, library science, health care, the humanities, just to name a few, have also taken an increasing interest in the Semantic Web. Large-scale industrial applications are under way or already deployed.

Yet, the division between computer science, natural science, and academia/government/industry, has a downside. It limits the formation of a coherent national agenda for exploring emerging trends in Semantic Technologies. What are needed are community consolidation and the building of a U.S.-based community research network.

The goal of the U.S. Semantic Technologies Symposium series is to bring together the U.S. Semantic Web community and begin forming such a research network. We achieve this by supporting communication across disciplinary, organizational, and geographical boundaries. The symposium events provide a forum by which participants can share information and ideas, coordinate ongoing or planned research activities, foster synthesis and new collaborations, develop community standards, and advance their science and education through communication and the sharing of ideas.

The first edition of US2TS[1] took place at Wright State University, Dayton, Ohio between March 1-2, 2018, and it served as the bootstrapping of the series. The first edition lined up a series of well known researchers and leaders in the Semantic Web field. Besides the invited talks, the symposium also offered open-discussion space through break-out sessions.

The second edition of US2TS[2] took place between March 11-13, 2019 at Duke University, Durham, North Carolina. The symposium attracted 120 participants from academia, industry and government. Following the feedback from the first edition, the 2019 sympo-

[1] http://us2ts.org/2018/
[2] http://us2ts.org/2019/

sium allowed the participants to co-create the program to include topics that were of interest to the community, and at the same time, it created a common discussion space through several plenary session.

2. The Symposium Program

The symposium ran over three days between March 11-13, 2019. The goal when creating the symposium program was to support the community in coming together not only around specific topics, but also around overarching themes that span different disciplines, organizations and geographical locations, and that affect the Semantic Technologies community as a whole.

The US2TS 2019 symposium program[3] was composed of two keynotes, 11 parallel sessions spanning the three days–each followed by a plenary report from the session,a poster session, two lightning talks sessions, and three additional plenary sessions: a panel on the "Semantic Web Technologies in the U.S. – five and ten years from now", a panel on "Common Challenges and Solutions in Using Semantic Technologies", and a town hall session.

The slides for several of the sessions are available online, and are linked from the US2TS 2019 web page.

2.1. The Keynotes

The symposium featured two inspiring keynotes given by Deborah McGuinness from the Rensselaer Polytechnic Institute, and by Helena Deus from Elsevier.

Deborah McGuiness gave a talk on "Knowledge Graphs Come of Age"[4] that revealed the diversity of use cases in which knowledge graphs are currently used, with a special focus on deployed applications in complex health care systems. She also discussed the current trends in knowledge graphs, and how they are gaining in popularity in artificial intelligence and data science applications.

Helena Deus gave a talk on "Building the Health Knowledge Graph: From Linked Data to Knowledge Graphs to Machine Leaning and back again".[5] The talk described the challenges and solutions that Elsevier encountered when using an expert-curated knowledge graph as the source of training data to do extraction of triples from medical literature. The talk ended with very practical suggestions on projects that the semantic technologies could help solve.

Both keynotes were well-received by the participants, and were often referenced throughout the symposium.

2.2. The Parallel Sessions

The 11 parallel sessions filled the main part of the symposium. We "crowd-sourced" from the participants these sessions ahead of the symposium. The structure of the symposium was intentionally designed bottom-up to enable the community to create a program that was highly relevant to the participants.

We gathered the request for discussion sessions from the community with an open call,[6] where anyone with an interest in Semantic Web technologies could propose a session. We defined session as any relevant activity that could fit into a 90-minutes slot, e.g. a panel discussion, a series of presentations on a topic, a breakout-style discussion on a proposed topic, a tutorial etc. The two main recommendations we gave to organize a session where to have at least 2-3 people as organizers, preferably from different institutions, and to cover cross-discipline topics as much as possible.

The response from the community was quite positive, with a total of 11 accepted sessions. The proposed sessions were of three main types:

1. *Foundational Sessions*: Addressing core topics of the field;
2. *Applied Session*: Addressing the application of semantic technologies to specific verticals;
3. *Practical Sessions*: Addressing the usage of semantic technologies in real business applications, and solutions to overcome the existing barriers.

2.2.1. Foundational Sessions: Definitions, Methodologies, and Algorithms

We classified under *foundational* those sessions that encouraged discussion and reflection on formal definitions and methodological problems related to semantic technologies.

The *Knowledge Graphs* (KG) session gathered different interpretations of what constitutes a Knowledge

[3] http://us2ts.org/2019/posts/program.html

[4] http://us2ts.org/2019/posts/program-session-keynote-deborah.html

[5] http://us2ts.org/2019/posts/program-session-keynote-helena.html

[6] http://us2ts.org/2019/posts/cfp.html

Graph, with the aim of coming to an agreed definition and sharing experiences on constructing large and small KGs in a variety of domains. A panel of experts from academia and industry shared their rationalization of the concept of KG. While coming up with a unique definition is a challenging task [1], the main message that surfaced is that one of the successes of KGs is data integration. It also became clear that KGs are the only viable way of dealing with large scale applications that require organizing and accessing data.

The *Identifying Cross-Domain Ontology Design Patterns* session focused on trying to define common modeling patterns that span different domains. The state of the art differs widely among different application domains. While some domains, such as the life sciences and some fields in industry, apply modeling patterns more commonly, patterns are rarely used in other domains. This session spanned two 90-minute slots and was organized as two break-out sessions, in which participants self-selected into two groups of common interests. The first group worked on several design patterns, such as the "agent-role" pattern and qualified relationships. The second group discussed patterns for modeling processes. The participants of the session agreed to continue the work through a mailing list. Another outcome is the report[7] by Chris Mungall who documented how Knowledge Graph modeling patterns may differ from the RDF/OWL patterns.

The breakout session *Towards Fusion of Semantic Knowledge into Deep Learning Models*[8] discussed how semantic technologies can help deep learning models to incorporate common sense about the real world and achieve sense making on a multi-model level, via injecting knowledge, i.e. ontologies, into the deep models.

The *Pushing the boundaries on reasoning applications to promote discovery* session addressed some of the big opportunities of different types of reasoning, but at the same time recognizing that the applications of formal reasoning are hampered by hard challenges and trade-offs, such as between logical expressivity and scaling reasoner performance, and a tool ecosystem with many gaps. This session brought awareness of both opportunities and challenges, and tried to coalesce the otherwise disparate community around tackling common gaps. Several well-received talks described real-world use cases of using reasoning, such as the development, maintenance and use of the Gene Ontology [2]; the challenges of reasoning over the semantics of shared descent with phylogenetic trees; the indispensability of reasoning in the development and maintenance of OBO Foundry ontologies [3]; or the state of OWL reasoning systems.

The tutorial on *the Role of Data Semantics for Explainable AI*[9] explored the current state of the art on explaining AI and how semantic technologies can help, proposing methods to shift the classification models towards hierarchical models that incorporate ontological models from the beginning.

2.2.2. *Applied Sessions: Semantic Technologies for Specific Verticals*

The session on *Rich Spatial Semantics* explored the topics of Locations, Places, and Spatial Relations and advocated for the need of richer and more nuanced geospatial concepts and vocabularies. The various talks examined strategies for improving and versioning available ontologies, as well as connecting formal and natural language representations of spatial knowledge.

The session on *Agriculture and Food* explored the application of semantic technology to the logistical, security and assurance problems of food and agricultural supply chains. One of the takeaway messages from the session was to focus on controlled vocabularies and on making them easy to use and adopt, then annotate them, i.e. match them to ontologies.

The *Traits, Phenotypes, Diseases, and Qualities* session addressed the need to come to an agreement on the representation of characteristics of entities and processes. These characteristics have been referred to as traits, phenotypes, diseases, and qualities, sometimes interchangeably or inconsistently. Each community has developed its own design patterns, classes, and properties for representing characteristics, sometimes in isolation. This session tried to bring clarity on why and how communities of practice are representing characteristics to avoid and remove unnecessary silos.

The *Using MediaWiki and WikiBase as a Platform for Library Linked Data* session described a recently completed pilot driven by the Online Computer Library Center (OCLC) using MediaWiki and WikiBase as a platform for creating and editing Linked Data. Their local installation was populated with linked-data entities mined from the OCLC WorldCat database,

[7] https://douroucouli.wordpress.com/2019/03/14/biological-knowledge-graph-modeling-design-patterns/

[8] https://deepsemantic2019.github.io/

[9] https://semanticsforxai.github.io/

which were synchronized with descriptions from corresponding Wikidata pages. The speakers of the session described the technical aspects of deploying the wiki for Link Data authoring, and also discussed lessons learned from their project.

2.2.3. Practical Sessions: Fostering Adoption of Semantic Technologies in Industry

The session on *Bringing Semantics to Enterprise Data* addressed the technical and social challenges to bring KG to the enterprise. The take-away message was about focusing on the end goal, i.e. what is the value that a KG adds? And how to communicate and expose the value to people in the organization who are not semantic technologies experts but rather UX designers, data scientists, etc.

One other aspect of making semantic technologies more appealing for industry was addressed in the session *Toward Easier RDF*, as the RDF ecosystem is perceived as too hard for developers. The session explored a few potential solutions, like focusing on property graphs as well as putting the focus on the "marriage" of RDF and JSON, and the success of GraphQL.[10] Another issue that was discussed was how to create a central webpage that could serve as a starting point for people wanting to build an application using semantic technologies.

2.3. Challenges in Using Semantic Technologies

One of the highlights of the symposium was the plenary session that took place on the third day on the topic of "Common Challenges and Solutions in Using Semantic Technologies". We asked the participants to submit the challenges they faced with semantic technologies through a Google form[11] during the symposium. We also asked six participants from diverse backgrounds to give a 5-minute position talk at the beginning of the session. Then, we opened the floor for discussions, and also invited all participants to edit a shared Google document[12] during the session. The discussions were very productive, identifying challenges that span across fields and use cases, and also generating ideas about ways to move forward. Below are some of the main ideas. The interested reader can find a complete reference in the shared Google document referenced above.

- *Modeling is hard*: Participants from different domains and with different expertise levels, ranging from novices to experts, raised several issues related to the difficulty of modeling of ontologies and KGs: There is a lack of good methodologies and modeling patterns for different use cases; it is very difficult to change a modeling pattern once you made a commitment; we need simpler modeling languages and better schema languages; there should be simpler ways to deal with constraint checking and reification. These issues are aggravated by a lack of tool support.
- *Need for education and professionalisation*: This issue found a lot of support from most participants. There is a substantial need for better education in using semantic technologies, and it should start in academic environments. Is it possible to develop a common curriculum? We need to develop a training programs that creates the Knowledge Engineer profession.
- *Lowering the barriers for entry*: This is definitely an old challenge, but obviously still unresolved, as many of the participants agreed. It is still very difficult for newcomers in the field to find the right resources to start developing with semantic technologies; there isn't a central webpage or hub with the most important resources. We also need better tools that are built with the user in mind. There is also a need for better ways of documenting ontologies to make them appropriate also for "human consumption".
- *Trust and maintainability*: Even though there are many ontologies and Linked datasets published on the Web, it is not clear which ones can be trusted. We need some kind of "stamp of approval" by accredited entities. Also licensing of ontologies and linked data is very important, especially for companies who would like to make use of them.

3. Feedback and Takeaway Messages

At the end of the symposium we run a town hall session to gather feedback from all participants on what worked well and what could be done better next year. The plenary and interactive sessions were well received, specifically the poster sessions and the discussion of the challenges. Also the opportunity of giving short lightening talks was well received. Sometimes

[10] https://graphql.org/
[11] https://goo.gl/forms/pREFgjT25jbOcqXH3
[12] http://tinyurl.com/US2TS2019

having three parallel sessions made it harder to participate in some of the discussions.

During the town hall discussion, two major improvements for next editions have emerged. Firstly, it was suggested to strongly clarify and share the goals of the symposium. While the primary goal of this event so far it has been to create awareness of each other within the US semantic technologies community, the goals going forward should become more concrete and content-based. Secondly, in the spirit of the community becoming more solid and established, more of the activities of the symposium should be proposed by the crowd during the event itself, rather than all the sessions being defined before the event. Essentially, the suggestion is to leave more space for interactive sessions, where participants can gather together during the event and decide what to discuss about.

The town hall session also revealed the overwhelming interests of the participants to have the third edition of the US2TS series organized next year rather than in two years. As of now, US2TS 2020 will be organized in Spring 2020 at the Woods Hole Oceanographic Institution, Massachusetts. Next year's edition will continue to promote the formation of a U.S.-based community research network around semantic technologies and it will be shaped by the feedback we received from the participants in this and the previous edition of the event.

4. Acknowledgements

We would like to thank the entire organization committee for making this symposium possible: Pascal Hitzler (General Chair), Hilmar Lapp (Local Chair), Marshall X Ma (Sponsorships), Amit Joshi (Publicity) and Krzysztof Janowicz (Outgoing Program Chair). We would also like to warmly thank our sponsors, listed at http://us2ts.org/2019/posts/sponsor.html, without which this event would not have been possible.

References

[1] P.A. Bonatti, S. Decker, A. Polleres and V. Presutti, Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371), *Dagstuhl Reports* **8**(9) (2019), 29–111. doi:10.4230/DagRep.8.9.29. http://drops.dagstuhl.de/opus/volltexte/2019/10328.

[2] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig et al., Gene ontology: tool for the unification of biology, *Nature genetics* **25**(1) (2000), 25.

[3] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall et al., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature biotechnology* **25**(11) (2007), 1251.