

Semantic Annotation of Collaborative Work Environments

Kenneth Baclawski*

April 15, 2018

Abstract

Environments for collaborative work and social networking are increasingly complex. There is a clear need for these environments to provide support for semantics. This report discusses work on semantic annotation and interoperability for such environments that began in a hackathon at the Ontology Summit 2014. Our hackathon was primarily concerned with developing semantic tools for the Ontolog Community Wiki, but applies to ecosystems that involve many tools and environments. The hackathon spawned subsequent work that is of independent interest, including the development of an ontology mapping tool for customizing generic repository software for more specific purposes.

Keywords: semantic annotation, semantic interoperability, collaborative work environments, social networking, MediaWiki

1 Introduction

The purpose of the Ontology Summit 2014 “Big Data and Semantic Web Meet Applied Ontology” was to build bridges between the Semantic Web, Linked Data, Big Data, and Applied Ontology communities. The Ontology Summit 2014 included a series of hackathons. This article reports on the results of the hackathon titled, “Adding semantic annotations and tools to Semantic MediaWiki websites for the Ontology Summit and the Ontolog Community Wiki” (Baclawski, 2014).

The article begins with some background about wikis in general, and the Ontolog Community Wiki in particular, as well as the standard for collaboration tools. Section 3 discusses the objectives of the hackathon; namely, to add explicit, accessible semantics to the Ontology Community Wiki so that it serves as a semantic interoperability ecosystem. Section 4 lists the hackathon participants and their contributions to the hackathon objectives. The hackathon efforts did not end when the hackathon ended, and these continuing efforts are described in Section 5. The article ends with a description of the plans for future work in Section 6.

*Northeastern University, Boston, MA USA; corresponding author email: kenbaclawski@gmail.com

2 Background

A wiki is a website that allows a user to modify the content and structure collaboratively directly from any web browser. To create a new page the user can simply request its URL like any other page of the wiki. The content for a wiki page is a simplified markup known as “wiki markup” or “wikitext.” Fig. 1 gives an example of what the markup looks like. The interpretation of the markup is explained in Section 2.1 below. The use of simplified markup has the advantage that very little specialized knowledge is necessary for creating and modifying wiki pages. A wiki is run by means of software called the wiki engine. There are dozens of wiki engines, and their wiki markup languages differ from one another. MediaWiki is the wiki engine for Wikipedia and is often regarded as the most popular wiki engine. Semantic MediaWiki (SMW) is a collection of plug-ins for MediaWiki that adds support for semantic annotations.

```
{| class="wikitable" style="float:right; margin-left: 10px;" border="1" cellpadding="10"
|-
! scope="row" | Session
| [[session::Postmortem]]
|-
! scope="row" | Duration
| [[duration::1.5 hour]]
|-
! scope="row" rowspan="3" | Date/Time
| [[has date::Jun 07 2017 16:30 GMT]]
|-
| 9:30am PDT/12:30pm EDT
|-
| 4:30pm GMT/5:30pm CET
|-
! scope="row" | Convener
| [[convener::KenBaclawski]]
|}

= [[OntologySummit2017|Ontology Summit 2017]] Post-Mortem =

== Abstract ==
In this session, we review the Summit process and make plans for the future.

== Agenda ==
* 1. Presentation of the Communiqu&eacute;:
** See [bit.ly/2uzIdof Communiqu&eacute;]
* 2. Summary of Ontology Summit 2017 (Summit Chair)
* 3. Summary of the Symposium (Symposium Chair)
* 4. Open Discussion
** What went well?
** What can be improved?
** Follow-up activities
* 5. Brainstorming on future Ontology Summit themes
** [[Blog:OntologySummit2018Themes|Blog for Ontology Summit 2018 Theme Suggestions]]

== Previous Meetings ==
{{#ask: [[Category:OntologySummit2017]] [[Category:Icom_conf_Conference]]
|[has date:<<{{#show:{{PAGENAME}}|?has date}}]] |?mainlabel=-|sort=has date
|order=desc| format=ull limit=3}{

[[Category:OntologySummit]]
[[Category:OntologySummit2017]]
```

Figure 1: Example of the markup of a wiki page

Purple Semantic MediaWiki (PSMW) is a plug-in to SMW that allows fine-grained

addressability to the content of wiki pages. PSMW achieves its goal of fine-grained addressability by adding identifiers called “purple numbers” at the end of content sections on each wiki page. Unlike most other kinds of web pages, the content on a wiki page can be the result of a dynamic collaboration among the users of the wiki. As a result, the content on a wiki changes more frequently than most web pages, and the ability to pinpoint the location of specific data on the page is important not only to link between pages but also to communicate between people who are collaboratively developing content (Baclawski, Parikh, Gupta, Yim, & Cheyer, 2010). Purple numbers implicitly reify the parts of a wiki page. Making this reification explicit with appropriate semantics is an interesting problem that is proposed as future work in Section 6 below.

The Ontolog Community Wiki has always served as the ontology community’s dynamic knowledge repository. The migration of Ontolog Community Wiki content (in particular, the entire body of knowledge from Ontology Summits) from the earlier wiki to SMW had the following goals:

1. Organize the Ontolog Community Wiki content in a way that encourages and facilitates access to and (re)use of the material
2. Support semantic annotations using standard ontologies
3. Develop a semantic ecosystem with new functionality and views of the content

The hackathon was part of a continuing effort to accomplish these goals.

2.1 Semantic MediaWiki

SMW is an extension of MediaWiki that helps to search, organize, tag, browse, query, and share a wiki’s content. SMW provides a semantic infrastructure for a wiki. From the point of view of an ontology, individuals (resources) are the wiki pages. An example of the markup for a wiki page is shown in Fig. 1. This wiki page is for one of the sessions of the Ontology Summit 2017 “AI, Learning, Reasoning, and Ontologies (2017)”. The URL of this wiki page is http://ontologforum.org/index.php/ConferenceCall_2017_06_07. The wiki page begins with an “attribute box” that specifies the main properties of the wiki page as a resource. The attribute box specifies the session, date, time, duration and convener attributes of the page. These are all properties of the wiki page. In terms of RDF, they are triples where the subject is the wiki page (or more precisely, the URL of the wiki page). When the wiki page is rendered, the attribute box appears as in Fig. 2. Properties of a wiki page can be specified anywhere in the page, but an attribute box is a useful way to display the properties of the page.

The main body of the wiki page uses ordinary MediaWiki markup. Headings are specified with equal signs used as paired delimiters, with the number of equal signs specifying the level of the heading. Lists are specified with asterisks, with the number of asterisks specifying the level of the list item. Double brackets specify links to other wiki pages. Single brackets specify external links. The name of a blog page is prefixed with “Blog.”

Session	Postmortem
Duration	1.5 hour
Date/Time	Jun 07 2017 16:30 GMT 9:30am PDT/12:30pm EDT 4:30pm GMT/5:30pm CET
Convener	KenBaclawski

Figure 2: Example of an attribute box displaying the SMW properties of a wiki page

The construct starting with “#ask” and enclosed in double braces is an SMW query. The SMW query language can be used to retrieve the links and properties of selected wiki pages. For example, it is useful on a session page to list the previous session pages in the series. Doing this manually is tedious and error-prone. The query in Fig. 1 lists the previous three sessions. The query is executed dynamically as part of the rendering of the wiki page. The query uses categories to restrict to sessions in the same series. In this case, the desired pages are the conferences in the OntologySummit2017 category that are chronologically prior to the current session, sorted in reverse order by date and limited to 3 session pages. The result of the query is shown on the wiki page as in Fig. 3. Clicking on “...further results” opens a page that lists more results of the query.

Previous Meetings	
	ConferenceCall 2017 05 10
	ConferenceCall 2017 05 03
	ConferenceCall 2017 04 26
	... further results

Figure 3: Example of the result of a SMW query

The last two lines of the wiki page markup specify the “categories” of this page. SMW categories are analogous to classes in an ontology. Each category has its own wiki page that defines the category and lists the wiki pages that are in this category. The name of a category page is prefixed with “Category:.” A user can create a new category by simply creating a page with the Category: prefix. A category page can also specify categories and these categories will be supercategories of the category. As a result, categories form a subcategory hierarchy that corresponds to the ontological notion of a subclass hierarchy. Both datatype properties and object properties are supported, but object properties can only link wiki pages to each other. The semantic content of all or part of an SMW wiki can be exported to RDF format so that the wiki can function within Semantic Web and Linked Data frameworks. The RDF export of the wiki page whose markup is shown in Fig. 1 is shown in Fig. 4.

```

<rdf:RDF>
  <owl:Ontology rdf:about="&wikiurl;Special:ExportRDF/ConferenceCall_2017_06_07">
    <swi:creationDate rdf:datatype="xsd:dateTime">2017-10-02T00:46:11-04:00</swi:creationDate>
    <owl:imports rdf:resource="http://semantic-mediawiki.org/swi1t/1.0"/>
  </owl:Ontology>
  <swi:Subject rdf:about="&wiki;ConferenceCall_2017_06_07">
    <rdf:type rdf:resource="&wiki;Category:300ontologySummit2017"/>
    <rdf:type rdf:resource="&wiki;Category:300icom_conf_Conference"/>
    <rdfs:label>ConferenceCall 2017 06 07</rdfs:label>
    <swi:page rdf:resource="&wikiurl;ConferenceCall_2017_06_07"/>
    <rdfs:isDefinedBy rdf:resource="&wikiurl;Special:ExportRDF/ConferenceCall_2017_06_07"/>
    <swi:wikiNamespace rdf:datatype="xsd:integer">0</swi:wikiNamespace>
    <property:Convenor rdf:resource="&wiki;KenBaclawski"/>
    <property:Duration rdf:datatype="xsd:double">1.5</property:Duration>
    <property:Has_date rdf:datatype="xsd:dateTime">2017-06-07T16:30:00Z</property:Has_date>
    <property:Has_date-23aux rdf:datatype="xsd:double">2457912.1875</property:Has_date-23aux>
    <property:Session rdf:resource="&wiki;Postmortem"/>
    <swi:wikiPageModificationDate rdf:datatype="xsd:dateTime">2017-08-10T21:34:26Z</swi:wikiPageModificationDate>
    <property:Modification_date-23aux rdf:datatype="xsd:double">2457976.398912</property:Modification_date-23aux>
    <swi:wikiPageSortKey rdf:datatype="xsd:string">ConferenceCall 2017 06 07</swi:wikiPageSortKey>
  </swi:Subject>
  ...
</rdf:RDF>

```

Figure 4: Example of the RDF representation of a wiki page

While SMW provides ontological features to MediaWiki, it has many limitations, even in the limited context of MediaWiki. For example, the query language is not very general and is difficult to use. For the most part, the limitations were the result of design decisions by the developers as well as the limitations of the MediaWiki infrastructure. In theory, once the semantic content has been exported to RDF, Semantic Web tools can be used for querying and inferencing. However, having done so, there is no simple way to transfer the results back to the wiki.

2.2 ICOM

The OASIS Integrated Collaboration Object Model (ICOM) is a standard for objects, classes, attributes (Baclawski et al., 2013; Dragan, Baclawski, Chan, Durusau, & Lee, 2013), relationships, constraints, and behavior, for an integrated and interoperable collaboration environment. The ICOM specification is intended to help integrate a broad range of objects to enable seamless transitions across collaboration activities. This enables applications to provide continuity of conversations across multiple channels. For example, applications can aggregate conversation threads in email with other conversations on the same topic in instant messaging, over the phone or via real-time conferencing, and in discussion threads in a community forum, wiki, weblog or micro blog. One of the objectives of the hackathon project was to map ICOM to PSMW. This was not accomplished during the hackathon period, but two of the participants eventually completed this effort. This is discussed in Section 5 below. The left-hand side of Fig. 7 shows one small part of the ICOM data model.

3 Semantic Interoperability Ecosystems

According to TopQuadrant, a *semantic ecosystem* is a modern information infrastructure that uses a rich web of metadata and services to:

- Add meaning to raw data artifacts;

- Make relationships between and within data sources explicit;
- Dynamically bring information together when and as needed; and
- Enable patterns to be discovered “Top Quadrant (2016)”.

The Ontology Summit 2016 was concerned with ecosystems that not only have these four capabilities but also have semantic interoperability “Framing the Conversation (2016)”. By introducing explicit, accessible semantics to a collaborative development system, the Ontolog Community Wiki is an example of a semantic interoperability ecosystem, with the added feature that it supports semantically integrated collaborative development by a community. Accordingly, it is reasonable to refer to this ecosystem as the Ontolog Community Semantic Ecosystem (OCSE). The OCSE is used for organizing a variety of activities on behalf of the ontology community, including recording meetings and conferences, developing artifacts such as communiqués, use cases, proposed standards, and project deliverables. The OCSE also has pages for community members, organizations, projects and blogs. Fig. 5 shows the part of the top of a blog page for the Ontology Summit 2017, and Fig. 6 shows part of the bottom of the same blog page which is a link to a blog article of the blog.

The screenshot shows a blog page with the following structure:

- Header:** Buttons for Blog, Discussion, Read, Edit, View history, Post new article, and a dropdown menu.
- Title:** Blog:Machine Learning for Ontology Development
- Content Area:**
 - Contents [hide]:** A section containing links to 1 Purpose, 1.1 Automated and Machine Learning for Knowledge Extraction, Ontology Development and Enhancement, 1.1.1 Organized by Gary Berg-Cross (Ontolog Board Member), 1.2 Machine Learning Methods, and 2 References.
 - Purpose:** A heading followed by a large block of text.
 - Automated and Machine Learning for Knowledge Extraction, Ontology Development and Enhancement:** A bolded heading.
 - Organized by Gary Berg-Cross (Ontolog Board Member):** A bolded heading.
 - Text Content:** A detailed paragraph about the challenges of using ontologies in knowledge bases and the role of machine learning in automating ontology development.

Figure 5: Example of the top of a blog page

The advantage of semantic annotations is the ability to automatically generate other representations such as RDF for SMW as shown in Section 4. This is especially important for addressing the barriers and challenges to the use and reuse of ontologies in the Semantic Web, Linked Data, Big Data, and Applied Ontology communities.

Maedche, Alexander, and Steffen Staab. "Discovering conceptual relations from text." Ecai. Vol. 321. No. 325
 automatic engineering of ontologies from text." Proceedings of the 12th international conference on software
 Maedche, Alexander. Ontology learning for the semantic web. Vol. 665. Springer Science & Business Media, 2012.
 Simperl, Elena , Tobias Buerger, Simon Hengl, Stephan Woelger, and Igor Popov. Ontocom: A reliable cost est
 Semantics: Science, Services and Agents on the World Wide Web, 16(0):1 – 16, 2012.
 Wong, Wilson, Wei Liu, and Mohammed Bennamoun. "Ontology learning from text: A look back and into the future."
 For information about how to use the Ontolog Community Blog see Blog.

Hybrid System

Reasoning With Neural Tensor Networks for Knowledge Base Completion by Socher, Chen, Manning and Ng
 → continue reading...

Posted by GaryBergCross (talk) on 31 December 2016 at 22:23 (GMT), no comments.

Category: Blog

Figure 6: Example of the bottom of a blog page showing a link to a blog article

4 Hackathon Activities

During the hackathon, each participant assumed a responsibility toward the overall objectives.

- Peter Yim developed the requirements for the project and answered questions about the structure and formatting conventions of the wiki. Peter was the original developer of the Ontolog Community Wiki, so his experience was invaluable.
- Tejas Parikh provided systems support for the project. Tejas also developed the software for migrating the wiki from PurpleWiki to PSMW.
- Sankalpa Kulkarni developed software for crawling the Ontology Community Wiki. There are several ways to specify links on MediaWiki pages so crawling is more complex than crawling ordinary web pages.
- Akshay Hathwar developed software for retrieving and modifying the raw wiki text of wiki pages. The process of rendering a MediaWiki page is complex with many stages and hooks at the different stages, allowing plugins to customize the rendered page. The wiki text of a page is not normally available in its raw form so a mechanism had to be programmed to obtain it.
- Gaurav Durgule developed software for extracting properties from a wiki page (day of the meeting, time of the meeting, title of the meeting, etc.)
- Srinivas Varadharajan developed software for formatting the properties of a wiki page using SMW properties and also using JSON data.
- Rohith Vallu developed software for integrating the wiki with the Google calendar web service code.

- Vivek Chouhan was the overall integrator of the various software components developed by the other participants.

The remaining objectives were concerned with mapping ICOM to PSMW. Two participants, Balkrishna Desai and Ravneet Popli began the work on this part of the project, but it quickly became clear that this was much too difficult for the short time period of the hackathon, so it was only started. The project continued for the following year and expanded to include some additional objectives. This work is discussed in the next section.

5 The ICOM Mapping Project

A wiki is an example of a content management system (CMS). The Content Management Interoperability Services (CMIS) standard defines a domain model, Web Services and Restful AtomPub bindings that can be used by applications to work with one or more content management repositories or systems (Müller & Baclawski, 2015). The CMIS interface is designed to be layered on top of existing content management systems and their programmatic interfaces. This includes wikis such as PSMW as well as many other repositories.

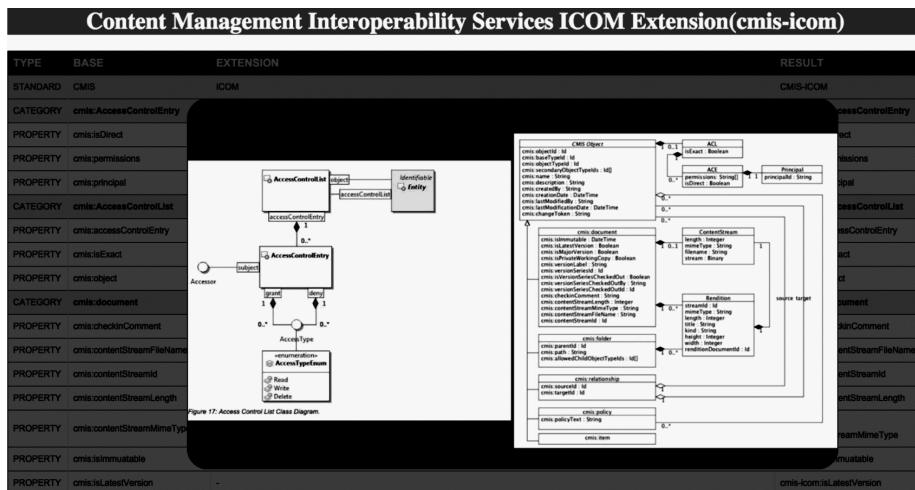


Figure 7: Example of the interface for the CMIS Generic Extender System showing the source ontologies being mapped

The CMIS domain model may be regarded as an ontology specified with a UML diagram as well as natural language descriptions of the axioms. The right side of Fig. 7 shows part of the CMIS ontology. The CMIS ontology is analogous to an “upper ontology,” such as the Basic Formal Ontology (BFO) (Arp, Smith, & Spear, 2015), or the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Gangemi, Guarino, Masolo, Oltramari, & Schneider, n.d.). The CMIS ontology has two levels,

the basic level shown in Fig. 7 as well as a metalevel. The purpose of the metalevel is to allow a repository to be customized with new document classes, document properties, and other kinds of classes and properties. If a customization is standardized, then the customization is analogous to a UML Profile “UML 2.0 Superstructure. (2005, page 633)”. Such a customization, when viewed as an ontology, is a domain-specific ontology that can be regarded as being “under” the CMIS “upper ontology.” Unlike most upper ontologies, CMIS specifies behavior in the form of Web Services as well as Restful AtomPub bindings so that any domain-specific ontology under CMIS would not require any additional implementation effort to be a web application.

The ICOM standard, on the other hand, defines a framework for integrating a broad range of domain models for collaboration activities in an integrated and interoperable environment. Like CMIS, ICOM has both a basic level and a metalevel. The ICOM metalevel allows an ICOM-compliant collaboration environment to be customized. Thus, ICOM may also be regarded as a “upper ontology” for collaboration environments. However, ICOM is more specific than CMIS, so it would be more accurate to regard ICOM as an intermediate ontology. In addition, unlike CMIS, ICOM does not specify the behavior of an ICOM-compliant system.

The purpose of the ICOM Mapping Project is to explicitly specify ICOM as a domain-specific ontology under CMIS. In other words, the purpose is to specify ICOM as a profile of CMIS, in the sense described above. Since both ICOM and CMIS are too large for this to be easily done entirely manually, a tool was developed to reduce the effort.

Amanda Priscilla Araujo da Silva developed an innovative tool called the CMIS Generic Extender System. This software supports the conversion of a given data model, specified by a similar standard such as ICOM, into an extension of CMIS. The system starts by mapping similar data between the CMIS data model and the model to be converted. Once the initial combined data model has been constructed, the user can manually complete the process by adding new mappings or editing the constructed mappings by using an interface that makes suggestions. While the tool was specific to CMIS, it could be modified for other ontologies. Sahil Chhabra focused on the details of ICOM and CMIS as ontologies, and developed the mapping between them using the CMIS Generic Extender System. An example of mappings that were found or manually entered are shown in Fig. 8.

6 Future Work

As noted in Section 2 above, purple numbers may be regarded as resource identifiers for parts of wiki pages. These resource identifiers can be used in Semantic Web and Linked Data contexts. In such contexts, a purple number resource is a part of the wiki page, not the wiki page as a whole. Consequently, the purple numbers implicitly reify the parts of a wiki page, and the parts implicitly form a “part of” (mereological) hierarchy. Unfortunately, the parts of a wiki page do not have the same status as wiki pages in SMW. For example, one cannot specify properties or categories for parts of a wiki page, and one cannot use such parts in an SMW query. Furthermore, the mereological hierarchy of a web page is also not queryable. For example, one cannot obtain the

TYPE	BASE	EXTENSION
STANDARD	CMIS	ICOM
CATEGORY	cmis:AccessControlEntry	icom_ac:AccessControlEntry
PROPERTY	cmis:isDirect	-
PROPERTY	cmis:permissions	-
PROPERTY	cmis:principal	-
CATEGORY	cmis:AccessControlList	icom_ac:AccessControlList
PROPERTY	cmis:accessControlEntry	icom_ac:accessControlEntry
PROPERTY	cmis:isExact	-

Figure 8: Example of the interface for the CMIS Generic Extender System

web page containing a given part of a web page, other than by performing a textual analysis of the URL of the purple number. We plan to develop the full mereological features, including equal status with wiki pages, explicit mereological structure, and inferencing.

We also hope to further develop the CMIS Generic Extender System, including the following features:

- Add a setup window where a user can indicate patterns for improving the automated matching of terms in the two ontologies.
- Use machine learning techniques to find matching patterns. This would reduce the effort required for developing patterns mentioned above.
- Add support for generating standardized HTML suitable for incorporation into a standards document. This would reduce the effort needed for standardizing a customization.
- Add an option for saving progress using a backup-recover strategy to avoid unnecessary work for the user and improve recoverability of the system. One possible way to accomplish this would be to apply the CMIS Generic Extender System to itself so that any standard CMIS-based repository could be employed for backup and recovery.
- Generalize the tool to support upper ontologies other than CMIS.

Acknowledgments

This article was published by IOS Press in *Applied Ontology*, 12(3-4), 313–322. The final publication is available at IOS Press through <http://dx.doi.org/10.3233/AO-170186>

References

- Arp, R., Smith, B., & Spear, A. (2015). *Building ontologies with Basic Formal Ontology*. The MIT Press.
- Baclawski, K. (2014). Adding semantic annotations and tools to Semantic MediaWiki websites for the Ontology Summit and the Ontolog Community Wiki. In *Ontology Summit 2014 Hackathon*.
- Baclawski, K., Chan, E., Dragan, L., Durusau, P., Lee, D., Yim, P., & Yin, Y. (2013). ICOM: A framework for integrated collaborative work environments. *WISE 2011 and 2012 Combined Workshops, LNCS*, 7652, 147–158.
- Baclawski, K., Parikh, T., Gupta, V., Yim, P., & Cheyer, J. (2010). Fine-grained access to online content for virtual communities. *Intern. J. Knowledge, Culture and Change Management*, 9(12), 109–126.
- Dragan, L., Baclawski, K., Chan, E., Durusau, P., & Lee, D. (2013). ICOM: A framework for integrated collaborative work environments, part 2. *java.net*.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (n.d.). *Sweetening ontologies with DOLCE*. Rome/Padua, Italy. (Retrieved July 19, 2017 from <http://bit.ly/2f0R4PV>)
- Müller, F., & Baclawski, K. (Eds.). (2015). *Content Management Interoperability Services (CMIS) Version 1.1 Plus Errata 01*. Organization for the Advancement of Structured Information. (Retrieved July 19, 2017 from <http://bit.ly/12o0eyP>)
- “Framing the Conversation: Ontologies within Semantic Interoperability Ecosystems. Ontology Summit. (2016). Retrieved July 19, 2017 from <http://bit.ly/2f0GwQW>”.
- “AI, Learning, Reasoning, and Ontologies. Ontology Summit. (2017). Retrieved July 19, 2017 from <http://bit.ly/2jMRWFo>”.
- “Top Quadrant website. (2016). Retrieved July 19, 2017 from <http://bit.ly/1ezhc30>”.
- “UML 2.0 Superstructure Specification. (2005). Retrieved July 19, 2017 from <http://bit.ly/2fPmki1>”.