Modeling Social Movement Theory for the Intelligence Community: a Case Study[1]

Line Pouchard, PhD[1], Patrick Paulson, PhD[2], Ryan Hohimer[2]

[1]Oak Ridge National Laboratory

[2]Pacific Northwest National Laboratory

Contact: pouchardlc@ornl.gov

Abstract. We describe the modeling of concepts taken from social and behavioral science theories, specifically Social Movement Theory, with OWL ontologies. One goal of the study was to provide analysts with a way to understand the potential contribution of social movement research to their tasks without interrupting workflow. Another goal was to facilitate knowledge transfer at the point of need between busy analysts and the latest research. Instead of providing mechanisms for obtaining consensus on debated concepts, we used different namespaces for each author and record bibliographic information in another ontology. We also describe a framework for knowledge discovery supported by the ontologies.

Introduction

The dissemination of social science knowledge has traditionally been achieved through published academic papers that may not be accessible to intelligence analysts, both because of availability or because they may not be trained as social scientists. Additional constraints are placed on these users by their lack of time and resources relative to the amount of Open Source Intelligence at their disposal, in particular when using the Web to search for information (Mercado 2004, Clark 2004). The volume and complexity of information that may be needed to usefully apply social science concepts to "real world" situations requires the assistance of computerized tools to help users in a variety of ways. In particular, analysts may engage in (1) investigating and representing complex networks as needed within a specific task (2) understanding the larger theoretical and empirical context that underpins certain concepts, (3) accessing detailed information about specific cases or events, (4) adding data to support theoretical models as new evidence emerges (Clark 2004).

Social Movement Theory (SMT) has recently become of interest to the intelligence community to understand the emergence of modern social movements, such as the rise of fundamentalism in the context of state policies. SMT studies ideas, people, events, and organizations that are linked together in collective action and attempts to provide a theoretical framework for the study of social movements (Della Porta 2006). This theory has been

---

[1] US Government sponsored the work described in this paper. Therefore, certain data had to be omitted due to the terms of the sponsorship. These omissions do not affect the technical content and expository clarity of the paper. The submitted manuscript has been authored in part by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

applied to several modern social movements that have emerged as a common feature of the political landscape, such as anti-globalization, environmentalism, and the anti-abortion movement. Recently, some specialists of Islamic activism have started to explore SMT to find cross-disciplinary methods and found fruitful areas of convergence (Wiktorowicz 2004). Access to institutionalized politics and political opportunities can influence the rise of social movements and their propensity to use violent contention by providing them with more or less favorable external conditions (Hafez 2004).

This paper presents a case study of how ontologies encoded in OWL-DL, a Semantic Web technology, specify social science concepts in view of understanding the emergence and current conditions of social movements within the context of Social Movement Theory. One goal of the study was to provide analysts with a way to understand the potential contribution of social movement research to their tasks without interrupting the workflow. Another goal was to facilitate knowledge transfer at the point of need between the latest research and analysts who may be novices. The SMT ontologies were to support the design of a computer system that would organize the theoretical knowledge into machine processable entities and eventually accommodate the data values within these entities. This paper focuses on the challenging task of classifying a fluent and flexible body of theoretical knowledge represented in natural language into a computerized framework capable of accommodating empirical data. We present a method for encoding this knowledge, resolving conflicts between definitions, and anchoring ontology constructs into a larger, robust framework.

A prototype using components previously developed for the Semantic Web and available through open source licensing is also described. Encoding social science theory and knowledge in OWL, the standardized Web Ontology Language, (Bechhofer et al. 2004) and the development of a system that uses off-the-shelf standards allow current and future users to interact with the ontology. OWL was chosen to model SMT because many reliable open-source software components now exist and will continue to be developed as OWL has reached stable, standardized semantics. The OWL-DL flavor of OWL is preferred as it is optimized for reasoning and knowledge modeling. Reliance upon open source components also ensures that current users may extend the ontologies with new concepts. Other social scientists may also reference ontology definitions on the Web and adapt the prototype to capture knowledge in other social science domains (although with ontologies specific to their domain).

Ontology Engineering for the Social Sciences[2]

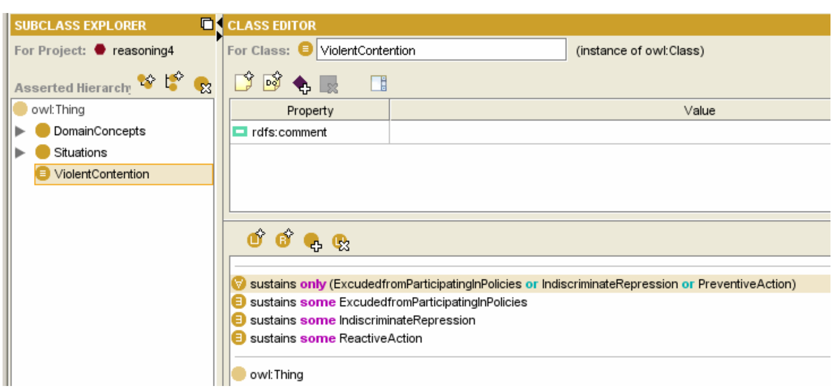

Figure 1: The rise of violent contention according to Hafez illustrated in OWL.

[2] Figures 1-3, 6-7, 9 were produced with Protégé, the ontology engineering tool.

SMT considers actors in a social movement as rational beings under structural constraints who are engaged in constructing cultural and framing processes, violence and contentious actions, and networks and alliances. Social and political organizations such as political parties or professional, charitable and religious organizations can be modeled as resources that are needed for supporting and coordinating collective action. For instance, the mosque in Muslim cultures often serves as a "religio-spatial mobilizing structure" (Wiktorowicz 2004). Figure 1 illustrates conditions for the rise of violent contention according to Hafez.

A defining aspect of SMT is its attempt to understand the behavior of groups within a society, situating itself between a macro- and micro level of analysis. The necessity to represent people in these ontologies (actors in the social movement, authors of particular theories that may disagree, etc.) as well as abstract concepts such as "theory" defined in natural language required a robust ontological infrastructure presented later in this section.

Using namespaces to resolve conflicting definitions for a given concept

OWL is designed to describe the semantics attached to resources. OWL entities are identifiers in some domain of discourse, they have properties and ontology rules that constrain their use. The concepts and relationships between concepts defining a particular theory are not fixed entities with unique and unambiguous semantics, as required for machine processing. They are subject to evolution over time and to nuancing or contradicting by social scientists. Given our users, ontologies for making claims about theoretical concepts was out of scope. An ontology for scholarly argumentation serving this purpose is described in Uren (2006). We have initially addressed the concept definition problem by specifying the usage context of the concept being defined. Within our OWL descriptions, this is accomplished by using a different entity for each definition of a particular concept. For example, Benford and Snow describe "collective action frames" as the products of "an active, processual phenomenon that implies agency and contention at the level of reality construction" (Benford 2000). Johnston (Johnston 2005) instead says "a collective action frame can be thought of as an aggregation of numerous individual interpretative schemata around an average." Our SMT ontologies contain both concepts, but in different *namespaces*, corresponding to different authors (Figures 2 and 3).

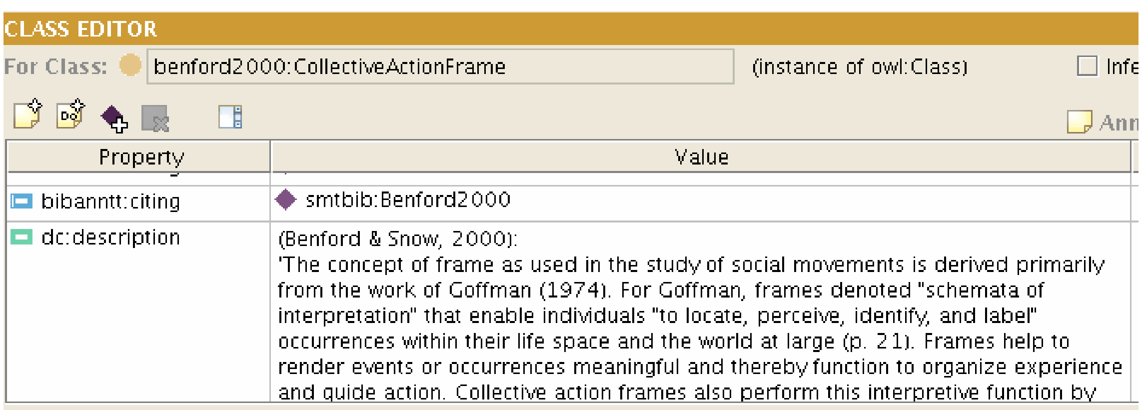

Figure 2 Namespace for Collective Action Frame according to Benford (2000): `benford2004b:CollectiveActionFrame.`

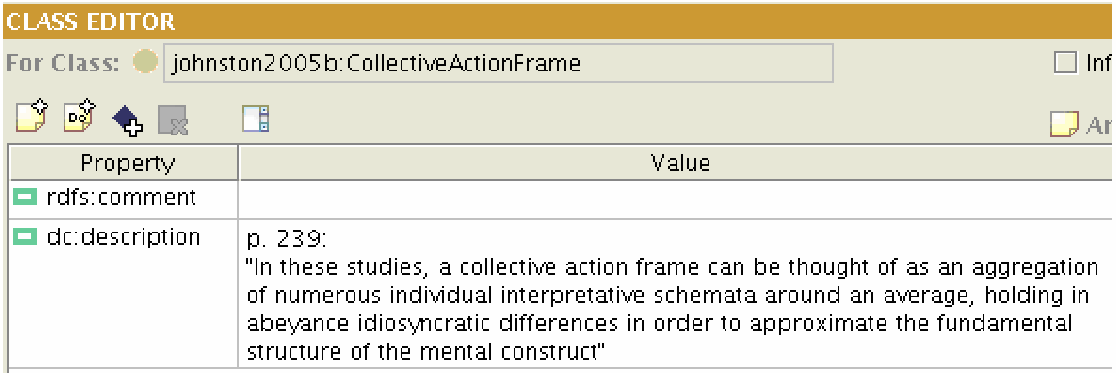

Figure 3. Namespace for Collective Action Frame according to Johnston (2005): `johnston2005b:CollectiveActionFrame`.

A namespace is a text string that is pre-pended to each member of a set of related concepts. In our case study, we used an abbreviation of the authors' names and a date, very much like references to citations in the body of a text. This design decision assumed that in each cited paper the author took care to use their terms consistently and did not contradict themselves in the definitions of concepts, at least within the space of each citation. In OWL and other Web languages, namespaces are defined by Universal Resource Identifiers (URIs) and unique for each resource:

> An XML namespace is identified by a URI reference; element and attribute names may be placed in an XML namespace using the mechanisms described in this specification. It is this combination of the universally managed URI namespace with the vocabulary's local names that is effective in avoiding name clashes (XML Core Working Group 2006).

This is usually achieved by using Universal Resource Locators such as Web addresses that describe how to locate each resource. For readability, namespaces within a specific system are generally abbreviated to a short string followed by a colon when no ambiguity will result. In our prototype system elements of the definition of frame is given several namespaces, corresponding to definitions by various authors

The decision to assign to each author its own namespace offered the double advantage of avoiding conflicts between definitions while providing a provenance trail. For instance, in our case study, an analyst not familiar with SMT concepts and their application to intelligence problems would be able to look up within the same framework the specific reference to a concept in the social science literature. She would also be able to follow the links to specific studies engaged in by an author and assess the applicability of a concept. Furthermore, the universal referentiality of namespaces on the Web provides the advantage potential re-use of a concept in other social science ontologies. If the ontologies of our case study were published on the Web, perhaps in a repository of social sciences ontologies, or on the web pages of this paper's authors, each concept within its namespace would acquire a unique address (a URL). It could then be pointed to by another Web ontology for reference. This URL may be of the type:

http://www.socialscienceontologies.org/benford2004b.owl#CollectiveActionFrame.

Achieving independence of entities and choosing the unit of analysis

Two entities are independent (disjointed) when they cannot share an individual (data value): each individual belongs to one or the other but not both. The independence of entities ensures that values are accurately classified by algorithms (for example, a person cannot be also a group or an organization). In choosing the unit of analysis, trade-offs are necessary with regard to the anticipated use of the ontologies: is representation of concepts at a high level of abstraction more suitable than a hierarchy with many levels? Representation at a high level uses less computer memory and the network of concepts and relationships is less complex. But a deep hierarchy will provide details that may be crucial to understanding nuances. In our case study, ontologies were designed to enable knowledge transfer to humans while remaining tractable in a computer environment for further processing.

Within the guiding principles of leveraging open source components, the choice of using high- level entities specified in Descriptive Ontology for Linguistic and Cognitive Engineering – Ultra-Lite (DOLCE-UL). DOLCE is an upper ontology concerned with the social reality and how it is created, known, accepted and modified by agents, often using natural language (Gangemi, 2002). For instance, in DOLCE UL, a Person is a Social Agent that may have a Description, Location, Role, participates in Event and may be a Member of a Collection (among other things). In the case study, we re-used some DOLCE-UL entities to define Theory and contextualize SMT ontologies. We propose that Theory may be defined as a unification criteria for collections of a certain type (Figure 4).

In Figure 4, the shaded circle "CommunityOfPractice" is an entity constructed for SMT ontologies by derivation (sub-class of) of the DUL entity "Collection." "Description," "Concept," and the properties "unifies" and "isDefinedBy" are also DUL properties.

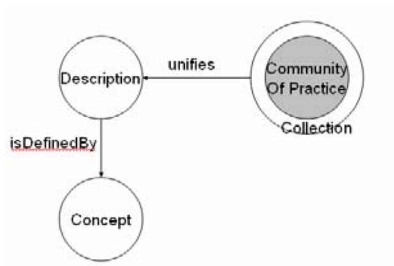

Figure 4. Useful DUL entities (nodes) and relationships (edges) to define Theory.

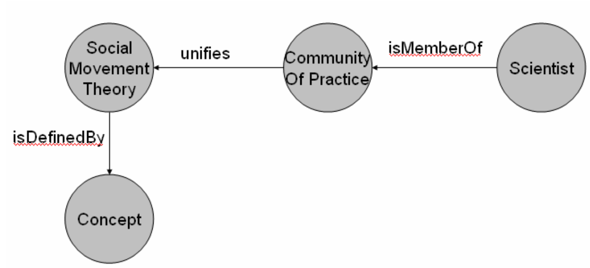

Figure 5. The derived entities defining Social Movement Theory.

SMT is defined as the description of an entity that unifies a certain group of social scientists. The DUL properties remain applicable. In our ontologies, Concept can serve as a top-level entity further specified with lower hierarchical levels, for example, resource mobilization, political opportunities and framing process. One benefit of this approach is the minimal amount of engineering spent on specifying entities that could answer the question, "what is a theory?" and still provide a theoretical context to support the development of SMT ontologies.

Another benefit is quality control of the SMT ontologies. Re-using general entities from an established body of knowledge not only enables the re-use of properties, but also enforces consistency in ontology design.

The proposed ontological definition of Theory directly supports the design decision made earlier to represent each author's work in its own namespace because this definition does not define concepts as a given truth, but as concepts organized in a description that unifies a community of practice. This is reflected in the use of namespaces where concepts may be defined as entities agreed upon by a certain group while still allow the possibility of disagreement as discussed in the previous section.

Many implications arise for social science from the description of a theory as a criterium of unification between a certain category of people. For instance, the name and number of practitioners of this theory provides an overview of its extent and popularity among a certain group of scientists. The unification property, if further defined, may indicate how closely certain researchers identify with a body of research. New work may be integrated by adding to scientist and/or concepts. Additional properties such as the applicability of concepts to specific situations may be derived from the DUL property between Concept and Situation, an important benefit for our case study, and for any concrete use of ontologies in a computer system.

Citing sources for terms in ontologies

In order for knowledge collections to be useful to social scientists, they must, like any source a scholar uses, transparently describe support for the assertions they claim. In scholarly articles this support is often provided by citation. While standards for providing citations for Semantic Web content are still emerging, our group has adopted the Dublin Core Metadata Initiative's vocabulary for information about a resource creator, the date when it was created and a brief text description (DCMI Usage Board 2005). While the Dublin Core Initiative has also proposed a method for documenting bibliographic information (Apps 2005), we have instead made use of an OWL adaptation of the BibTex bibliographic database system (Knouf 2003). BibTex has been widely used in academic publications since its introduction in 1985; many online bibliographies provide citations in the BibTex format (Wikipedia 2006).

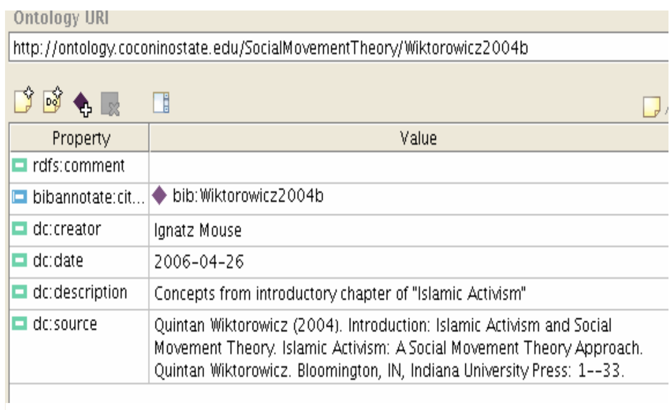

Figure 6. Source information attached to namespace.

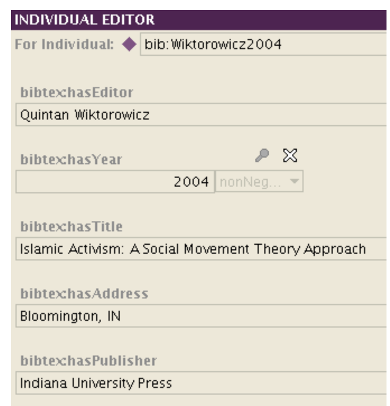

Figure 7. Publication details.

Figure 6 illustrates the documentation that is attached to the namespace 'wiktorowicz2004b'. Dublin core metadata properties (in the namespace 'dc') record that this ontology was created on 2006-04-09 by Ignatz Mouse, give a brief description of the ontology, and provide a citation for the source of terms in the ontology. The 'ontannotate:citing' link can be followed to find additional publication details given using properties derived from BibTex fields; an example is given in Figure 7.

This mechanism is useful to tie a particular concept or property to the literature; however, often concepts are described in terms of other concepts and properties. Documenting the relationship between multiple classes and properties, and tying that relationship to the literature requires a richer approach. Our current methodology is to associate a free-standing text document with each ontology. This text document follows the narrative of the source document and explains how concepts in the ontology are related to the source text, using hyperlinks to tie text to the ontology documentation.

Evaluation

Sure (2004) proposes that ontologies should be evaluated from three perspectives: a technology-focused evaluation, a user-focused evaluation, and a formal evaluation. Consistency and correct syntax (technology-based) were achieved by using an ontology engineering tool and grounding our work in an upper-level ontology.

Although the use of formal knowledge representation languages is sometimes limiting and non-intuitive, we have found that the precision formal modeling requires leads to a deeper understanding of the knowledge being encoded and to the various levels of interpretations made by social science theorists. Reasoning not directly supported by OWL inference engines may require specialized code in a final application.

The other two evaluation methods have not been used at this point due to the constraints of the project. In particular, the authors of this paper did not have access to analysts for a methodical evaluation within the scope of this study. However, credibility was achieved in two ways: by attaching machine-readable bibliographic information to each represented concept and by iterative knowledge elicitation from the authors when possible.

A Framework for Knowledge Discovery

In this section we describe a prototype system that allows analysts to access social science theories described by ontologies and use those theories to guide their analysis of documents and other sources. We assume that the analyst is working on a particular problem scenario and describe how the system helps the analyst apply social movement theory to the scenario.

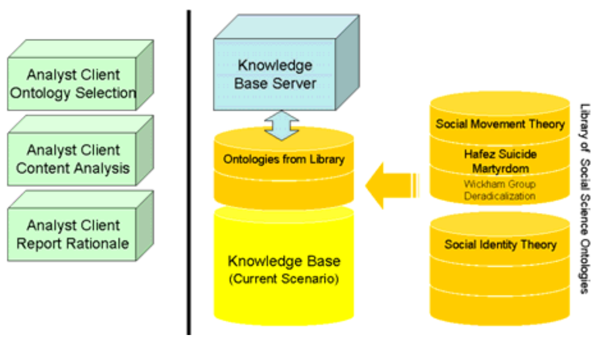

Figure 8. A Knowledge Discovery Framework.

System Design

Figure 8 illustrates the design of the system. Each ontology describing a particular social science theory is stored in a library of social science ontologies. These ontologies are accessed through a knowledge-server, which also provides access to a working knowledge base which contains ontologies and data associated with the current problem scenario. The analyst interacts with the knowledge base server through a graphical user interface (GUI), which provides functions for ontology selection, content analysis, and report generation.

Ontology Selection

The ontology library is indexed to allow discovery of the social science theories that are applicable to the analyst's current problem scenario. For example, searching the library for information about violent splinter groups would return ontologies based on (Hafez and

Wiktorowicz 2004); searching for information on the possible motivations for a possible suicide bombing would return ontologies based on (Bloom 2005).

Content Analysis

The content analysis function of the GUI provides the core strength of the analytical framework: it allows the analyst to view their problem scenario from the context of a particular social science theory. Through the content analysis function, the analyst examines the underlying ontologies and populates the knowledge base with data that matches the concepts provided in the ontology. The knowledge base is populated using background information, specific knowledge, and source documents. The user may also capture accumulated knowledge in the form of instances of news media that support or disprove a particular assessment or conclusion. Thanks to the digital encoding of major SMT concepts, analysts are able to look up online the information presented by social scientists within their coherent theoretical context and quickly find pointers to additional literature when more in-depth study is needed. In addition, the analyst may also organize his or her insights and record them into a theoretical framework in digital form in the knowledge base.

By populating the knowledge base, the analyst is able to determine which concepts from the underlying ontology should be considered in the scenario they are examining. Since the ontology provides not only the concepts and relationships used by the underlying theory, but also knowledge about the relationships between instances of specific concepts, the analyst can derive conclusions that are based on the synthesis of the data from their specific problem scenario and the accumulated knowledge of social science experts. For example, while examining press releases from a social movement organizer, an analyst might recognize an instance of “Boundary Framing” given the description from (Benford 2000). The developed ontology reminds the analyst that boundary framing is an indication of “Diagnostic Framing,” leading the analyst to look for further indications of Diagnostic Framing, including, perhaps, a party that the group may intend to target (Figure 9).

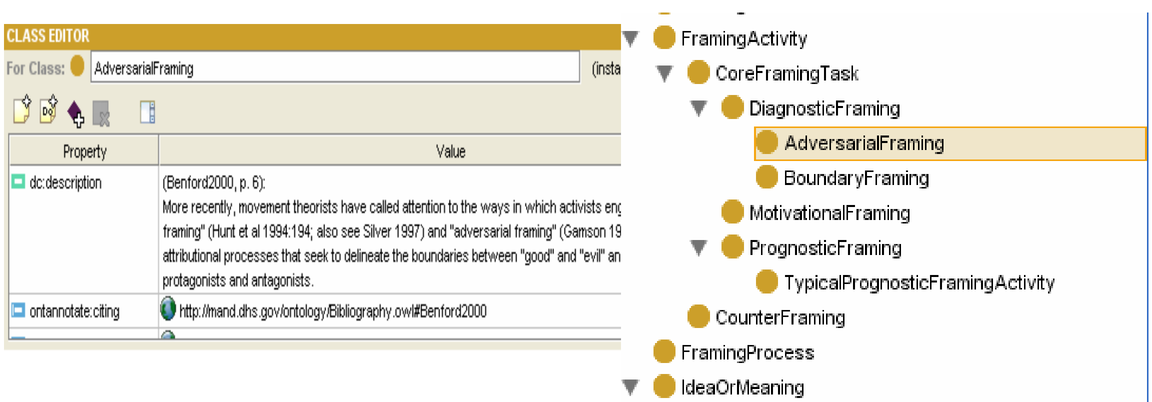

Figure 9. Recognizing diagnostic framing.

Automated Content Analysis

By attaching additional information to the concepts in the ontology, the capture of knowledge and constructs embedded in natural language can be automated. A methodology that extracts individuals of domain concepts from natural-language documents is described in Sanfilippo (2007). The method uses a combination of Information Extraction and Content Analysis techniques to find individuals (data values) that embody concepts described in an ontology In that paper, the authors have introduced an ontology for framing processes where the entities are Frame Promoter, Frame Issue, and Frame Target.

Report Rationale

Such a system can also be designed to assist the analyst in preparing final reports. The data instances collected by the analyst to populate the knowledge base form the evidence that a scenario matches a particular social science theory. The system not only has access to the evidence that the analyst used to draw conclusions, but also to a description of the theory the analyst chose along with references to the social science literature. The system could also be designed to keep a history of the analyst's actions to report on theories that were considered but rejected by the analyst.

Conclusion

Some studies of social movements rely on data acquired by mining text corpora produced by the participants in the movements that may be available on the Web. The information revolution offers new opportunities to study a social movement from its birth to its ultimate success or failure. However, without some automation, the amount of information may be too overwhelming for a single person to process. We have shown that ontologies could be beneficially used in social sciences in two aspects: a) for referencing concepts within their context and applicability, and b) at the backend of a system that allows for data analysis. Further work needs to be done to evaluate the ontologies with analysts.

References

Apps, A. (2005): Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata, Dublin Core Metadata Initiative.

S. Bechhofer, F. v. Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, B. L. R. Peter F. Patel-Schneider, and L. A. Stein (2004): "OWL Web Ontology Language Reference," World Wide Web Consortium, Recommendation.

Benford, R. D. and D. A. Snow (2000): "Framing processes and social movements: An overview and assessment." *Annual Review of Sociology* 26: 11-39.

Bloom, M. (2005): *Dying To Kill*. New York, Columbia University Press.

Clark, R. M. (2004): Intelligence Analysis. A Target-centric Approach. CQ Press, Washington, DC.

DCMI Usage Board (2005): DCMI Metadata Terms, Dublin Core Metadata Initiative.

Della Porta, D. and M. Diani (2006): *Social Movements: An Introduction.* 2nd edition. Blackwell, Oxford.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, and Schneider, A. (2002): "Sweetening Ontologies with DOLCE L." In A. Gómez-Pérez, V.R. Benjamins (eds.) *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, 13th International Conference, EKAW 2002, Siguenza, Spain, October 1-4.,

Hafez, M. M. and Q. Wiktorowicz (2004): "Violence as Contention in the Egyptian Islamic Movement," in Q. Wiktorowicz, (ed.) *Islamic Activism: A Social Movement Approach*, Bloomington, IN, Indiana University Press**:** 61-88.

Hayes, P. (2004): RDF Semantics, World Wide Web Consortium.

Johnston, H. (2005): Comparative Frame Analysis. *Frames of Protest: Social Movements and the Framing Perspective*. H. Johnston and J. A. Noakes. Lanham, MD, Rowman & Littlefield**:** 237-260.

Knouf, N. (2003): "Transformation of bibTeX into an OWL ontology," http://zeitkunst.org/bibtex/0.1/bibtex.owl.

Mercado, S. (2004): A Venerable Source in a New Era: Sailing the Sea of OSINT in the Information Age,' *Studies in Intelligence*, vol. 48, no. 3.

Sanfilippo, A., Cowell, A.J., Tratz, S.C., Boek, A.M., Cowell, A.K., Posse, C., and L. Pouchard (2007): "Content Analysis for Proactive Intelligence: Marshaling Fame Evidence, " *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, Vancouver, Canada, July 22-26, AAAI Press, Menlo Park, CA.

Sure, Y., Staab, S., and Studer, R. (2004): 'On-To-Knowledge Methodology (OTKM)', in S. Staab and R. Studer (eds.): *Handbook of Ontologies*, Springer, Heidelberg and New York.

Uren, Uren, V., Buckingham Shum, S.J., Li, G. and Bachler, M. (2006): "Sensemaking Tools for Understanding Research Literatures: Design, Implementation and User Evaluation," International Journal of Human Computer Studies, 64, (5), pp.420-445.

Wikipedia (2006): BibTex, Wikipedia, The Free Encyclopedia.

Wiktorowicz, Q. (2004): "Islamic Activism and Social Movement Theory," in *Islamic Activism: A Social Movement Approach*, Q. Wiktorowicz, Ed., Bloomington: Indiana University Press, 2004, pp. 1-33.

XML Core Working Group, (2006): "Namespaces in XML 1.0 (Second Edition)," World Wide Web Consortium Recommendation (16 August).