

Semantic Enrichment and Reasoning for Mobile Data Collection of Socio-Cultural Data

Alper Caglayan, Laura Cassani, Tim Clark, Jose Alavedra, Vinay Bharadwaj, Nick Lee, John Parkes

Milcord
303 Wyman Street, Suite 300
Waltham, MA 02169, USA

ABSTRACT

Leveraging the widespread adoption of mobile devices, we have developed a semantic knowledge management, mobile data collection, and situational awareness capability that enables the automatic semantic annotation and fusion of mobile collected data. The major components of our semantic knowledge management system consist of form management; data collection, aggregation, and storage; and spatially aware semantic analysis. Our research transforms the analysis of large amounts of data collected from mobile devices from an unstructured process to a structured approach leveraging semantic technology and automated tagging. Semantic technology is used to author, publish, and distribute forms to mobile devices. When submitted from the mobile device, the structured form is mapped to a semantic form to automatically generate semantic annotations from the mobile collected data for semantic enrichment. The semantic representation of the data enables users to organize, analyze, visualize, and create custom information products from the aggregated data.

Keywords: Semantic Knowledge Management, Geospatial Analysis, Mobile Data Collection, Socio-Cultural Data

INTRODUCTION

Estimates suggest that there are now more than five billion mobile phone subscribers in the world, with this trend increasing in relevance particularly in the developing world (Melanson, 2011). In support of the widespread adoption of mobile phones, we have developed a semantic knowledge management cloud service and mobile data collection and situational awareness capability called *SemLayer* that enables users to automatically semantically enrich mobile collected data, fusing information with contextual knowledge for analysis, visualization, and sharing purposes. Semantic enrichment is the process of augmenting unstructured data with semantic relations, defining concepts and the relationships between entities in an ontology. The benefits of harnessing semantic technology for information fusion and decision support are well represented in the literature (Caglayan, 2013; Caglayan et al, 2012; Stroh et al, 2010; Uren et al, 2006). By semantically enriching data, we are able to support semantic queries, allowing for analysis across multiple dimensions within the data. Semantic queries specify some conditions that describe what is asked for, such as the subject, category, properties, and values. Semantically enriched data also enables semantic reasoners to generate new relationships from the existing data (Horridge et al, 2008).

In support of this effort, we developed an integrated solution using a semantic wiki for knowledge management and an Open Data Kit (ODK) inspired mobile data management system. Semantic technology is used to author, publish and transmit structured forms to mobile devices for data collection. When submitted from the mobile device, the structured form is automatically mapped to a semantic form to generate semantic annotations from the mobile collected data for semantic enrichment. The semantic representation of the data enables users to organize, analyze, and create custom information products using semantic queries from the aggregated data, visualizing results with charts, tables, calendars, maps and timelines. We have also extended the semantic knowledge management constructs to represent and reason using geospatial methods. Mobile users can receive geospatially-enabled notifications and alerts to improve situational awareness and to refine the data collection plan.

We use semantic annotations as a semantic enrichment representation for knowledge management, leveraging machine automation to capture semantic enrichment instances such as form based data collection to capture semantic annotations; user defined semantic annotations in the wiki text; semantic inferencing to infer semantic enrichments from collected data; and semantic forms in the wiki to capture semantic annotations. The knowledge base automatically links categories, properties, and values in the data collected with contextual knowledge captured in the knowledge base ontology ingested from other data repositories. Twitter style markup enables the knowledge base to link user defined tags with existing categories, properties, and values, or creates these as new categories, properties, or values if they don't currently exist in the ontology.

Using a Description Logics (DL) formalism, semantic fusion consists of semantic queries that search across different category and property dimensions, with the semantic query results available in a variety of visualizations. Semantic queries fuse data across different data sources (contextual knowledge, mobile collected data, user defined data, inferred data, etc.), across different data dimensions (e.g. class (i.e. meta-data, schema) vs. instance (i.e. data)), and across different dimension instances (e.g. categories, properties). Using this format, we are able to support Level 2 and Level 3 data fusion by representing the operational environment and the relationships among the entities and the events in it for situation assessment and through semantic reasoning we are able to infer about the impact of the assessed situation and various vulnerabilities for the threat assessment level (Das, 2009).

Additionally, a DL based semantic solution is scalable to high density data streams. For example, several commercial solutions use triple stores that have been deployed to handle billions of facts. A DL based semantic solution is also flexible in that it is able to handle all types of data by incorporating the knowledge into a common ontology within the semantic store.

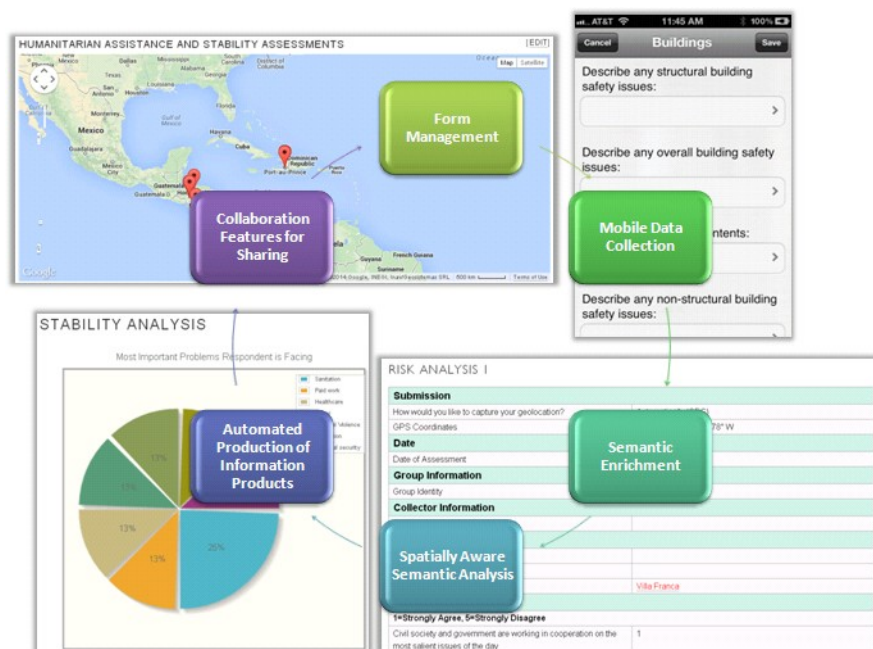


Figure 1. Semantic enrichment and fusion for mobile data collection conceptual framework.

SEMANTIC KNOWLEDGE REPRESENTATION

Description Logics

The underlying logical foundation for our research to leverage machine automation to generate semantic annotations as a semantic enrichment representation for knowledge management is Description Logics (DL), which are subsets of first-order logic. DL separates knowledge representation into: Terminological Box (TBox) for representing ontology classes and axioms about classes and Assertional Box (ABox) for describing specific objects and facts about individual objects. The TBox represents ontology classes and attributes, the hierarchy of concepts and the relationships between concepts. The ABox describes specific entities and facts about individual entities (Nardi and Brachman, 2003). The combination of TBox and ABox defines the full ontology, which is commonly referred to as the knowledge base. The DL representation enables semantic fusion which consists of semantic queries to search across different category and property dimensions.

We transform the structured knowledge collected on forms in mobile devices into semantic annotations using Resource Description Framework (RDF) triples. Formally, an annotation is a tuple consisting of the format “subject, property, object”. RDF and its extension OWL (Web Ontology Language) together provide an explicit and intuitive representation of common knowledge representation syntax for data transfer in the form of RDF triples.

Semantic Wikis

We use the Semantic MediaWiki platform as the basis of our knowledge management system. A semantic wiki differs from a regular wiki by enabling users to annotate wiki content by defining domain concepts, and relations between concepts, in a manner that can be queried like a database. Each wiki page represents an entity (e.g. person, organization, event, location) with the semantic wiki organizing the knowledge into categories and properties so users can easily search and discover relationships within the data. In contrast to a traditional wiki where thousands of users curate knowledge, querying the underlying structured knowledge base dynamically generates most of the pages in a semantic wiki.

Below we present a high level overview of semantic knowledge representation. Description Logics generates a hierarchy of concepts through ‘roles’. Such a hierarchy enables the children nodes to inherit the parent relationships. For instance, Boston has a population property because city has a population and Boston is an instance of the category city.

Table 1: Semantic knowledge representation

Knowledge Representation	Description Logics	Semantic MediaWiki	Example
TBox	Concept	Category	City
ABox	Individual Name	Is a	Boston
Relations	Role	Property	City [[has::population]]
Values		Object	Boston [[has population::625,000]]

Semantic wikis are the ideal platform for knowledge management because of the ontological organization of knowledge, the discovery of knowledge using faceted and semantic search, and linked data storage. For instance, adding semantics to wikis enables semantic search and navigation, provides a more intuitive interface, supports tagging and folksonomy management, and links the wiki content to external resources (El Ghali et al, 2007). There are also a number of benefits semantic wikis provide to support the learning process (Schaffert et al, 2006). For example, semantic annotations lead to reflection about knowledge. The analyst reflects on the content while performing a faceted search on the material yielding to the discovery of relations in the topic of interest. Second, semantic wikis enable fellow analysts to share formal models and build a common model collaboratively. Finally, reasoning and inference capabilities of semantic technologies can lead to discovery of knowledge without active user search thanks to the machine maintenance of organized knowledge (Horridge et al, 2008). The structured format of the semantic annotations enables the automatic generation of knowledge through dynamic tables, charts, graphs, and maps for analysis and visualization.

Semantic Forms

One of the mechanisms we use to capture the semantic annotations easily within the wiki is semantic forms, which enables the rapid input of structured content. Semantic forms are part of an extension to MediaWiki that allows users to use forms for adding, editing, and querying data without requiring any programming. Semantic forms enforce the use of templates by enabling a user to populate a pre-defined set of templates for a page. The data is turned into semantic properties once the page is saved. Forms can be used to edit the data in an existing page, with an “edit with form” tab showing on the top of any page for easy access. Semantic forms also support auto-completion of fields, so users can see previously entered values for a given field. This greatly helps to avoid issues of naming ambiguity and spelling errors.

MOBILE DATA COLLECTION

Semantic Authoring, Publishing, and Distribution of Forms

With an understanding of semantic knowledge representation concepts in our semantic wiki knowledge base, we turn to our mobile data collection capabilities that implements form based structured data collection for semantic annotation in support of semantic enrichment. To semantically enrich and fuse mobile collected data, we have developed an automated system using semantic technology to author, publish, and distribute structured forms to mobile devices and the semantic wiki knowledge base. We built an app that allows users to download forms to their mobile device, and fill out the form with the submissions automatically creating pages for each form instance within the semantic wiki. As the forms are ingested into the semantic wiki, the form fields map onto semantic forms so that data is automatically semantically annotated in the “subject, property, object” format, where subject is an instance of a category. By tagging data in this format, we are able to semantically fuse the data through the generation of semantic queries.

Every data collection form that exists within the mobile app resides as an Excel spreadsheet, or XLS Form, on the semantic wiki. These XLS forms adhere to a standardized format that allows them to be converted to XForms upon upload to the wiki. XForms are XML files that provide a data processing model for XML data and specify how the forms are presented to the user. The XForm contains the structure and content of the form; the full list of questions and information about how these questions should be presented. Most rows represent a question; the rest of the rows specify control structures such as groups. The spreadsheet also is used to specify the options for multiple choice questions, which helps to control the ontology. For example, mobile data collectors interacting with the public can record the sentiment of an interviewee using positive, negative, or neutral annotations, as opposed to having an open text field. Having the data collected in a structured format such as this enables a semantic query to aggregate data across dimensions such as understanding demographic data for those individuals that reported a negative view.

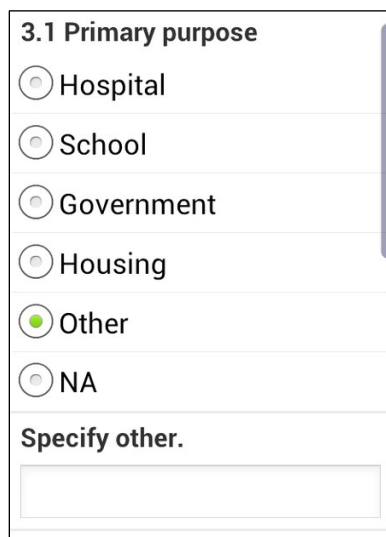
Users are able to download the XLS Form from the wiki, apply any necessary changes to the fields, and re-upload the spreadsheet into the wiki. Upon uploading the spreadsheet, the newest version of the form is represented with the automatic propagation of changes to the form, template, and XForm pages within the semantic wiki, as well as to forms within the mobile devices. This simplifies form management, as users are able to automatically update the mobile and semantic forms simultaneously, editing the ontology, while creating the necessary categories, properties and templates in the background, empowering real-time oversight of data collection requirements.

To demonstrate the types of properties associated with a particular category, consider the below example of a form that can be used for conducting a context analysis assessment. Here we can see the example of how the XLS Form specifies the field names, property names, and datatypes for semantically representing the data within the knowledge base. The field name is how the question appears in the form to the user on the mobile app and within the semantic wiki semantic form. The property name is the actual name of the property from the ontology. The datatype documents the property’s datatype, allowing the semantic wiki to properly evaluate the property value.

Table 2: Context analysis form field names, property names, and datatypes

Field Name	Property Name	Datatype
Identify the primary drivers of the conflict using the list of multiple selections	Primary driver of conflict	String
Hazards in the region have been nominal or non-existent	Level of hazard risk in the region	Number
Survey start date and time	Date and time of report	Date
List the NGOs operating in the area	Has NGO	Page
Primary purpose of the building	Has purpose	String
GPS Coordinates	Coordinates	Geographic coordinates

As the spreadsheet with the above outlined field names, property names, and datatypes is uploaded to the wiki, the form will be automatically distributed to the mobile app and the semantic wiki. Below we can see an example of how the form appears within the mobile app for data collection.



3.1 Primary purpose

Hospital

School

Government

Housing

Other

NA

Specify other.

Figure 2. Mobile app with question from context analysis form.

Semantic Annotations

As data is collected and the form is submitted from the mobile device, it automatically maps onto a semantic form with the associated semantically enriched data available in the semantic wiki. Each of the fields of the form are represented as standardized properties to create a common ontology. The specific assessment or form that is being completed becomes the subject, with the field represented as the property, and the inputted data as the object. As an example of a semantic annotation from mobile collected data represented in the “subject, property, object” format, the semantic wiki would store information about a particular building assessment using the name of the building unit as the subject such as “Clark Elementary, Has purpose, School” or “Clark Elementary, Year built, 1984”. Since each individual assessment is captured in a form on an individual wiki page, a user can go in manually and edit the semantic form at any time to add to or correct the field collected data. Using the mobile device’s GPS locator, we also automatically capture the coordinates of where the assessment took place to enable geospatial analysis and visualization. Below we can see an example of the semantic form within the semantic wiki.

Submission

How would you like to capture your geolocation? No answer Automatically (GPS) Manually (MGRS/Lat Long)

GPS Coordinates

Date

Date of Assessment

Group Information

Group Identity No answer Volunteer Community Staff

Collector Information

Assessment Collector

Peace Building and Conflict Sensitivity

What are the key historical turning points (international, national, and NGO sector) that have shaped the current conflict?

Identify the major actor-groups (4 maximum) who currently exert a positive influence on the conflict?

1)

2)

3)

4)

Figure 3. Context assessment form in the semantic wiki.

We present a summary of all of the queryable annotations in a factbox at the bottom of each page. Below we can see an example of some of the annotations captured in a context analysis assessment form submission.

Facts about Context Analysis 3	
Accelerators of conflict	Degraded local water source has exacerbated the social conflict +
Action ensuring that there is no discrimination	Yes +
Assessment team	Community +
Community capability rating	Excellent +
Coordinates	14.11262° N, 87.23308° W +
Country has National Disaster Management systems	No +
Country has disaster threats	Flooding + and Tsunami +
Country in disaster prone area	Yes +
Criminal groups view INGOs as desirable target	Yes +
Date / time of report	7 May 2013 +
Description of violent crime circumstances	Violent crime is specifically targeted at American aid workers operating within five miles of the village. +
Essential services infrastructure capability rating	Average +
Hazards caused loss of life	Yes +
Healthcare capability rating	Below Average +
Hostility to INGOs in country	Yes +
How capture geolocation	Manually (MGRS/Lat Long) +
Human induced vulnerabilities increase risk	Yes +
Intended action occurring to advance humanitarian imperative	No +
Local security forces sufficient and free of corruption	No +
Political violence exist in country	Yes +
Primary drivers of causes of conflict	Social + and Environmental +

Figure 4. Context assessment form factbox presenting a summary of semantic annotations.

Semantic Internal Objects

The advantages of employing a spreadsheet-driven solution to form authoring are manifold. In particular the structure of the XLS Form allows us to add augmented functionality. For example, our solution allows users to place groups of properties and their values inside semantic internal objects within the wiki. Semantic internal objects allow us to create “bundled annotations” which represent a collection of multiple values in a single annotation on a page. The value of leveraging these internal objects can be seen in the following example. Suppose a question within a form asks the user, “List this person’s siblings and their dates of birth” It is important to create distinct instances for each individual person and associate their birth date with the correct individual, if we are to treat each person as a unique instance and query it within semantic search. In this case, if we are to add multiple instances of people, we want to ensure that each additional property instance such as birth date is associated with the correct individual, as opposed to being associated with the overall form. We therefore use subobjects to represent these individuals within the factbox, and maintain their annotations separately. Leveraging semantic internal objects allows us to semantically store and query subgroups of compound values.

MOBILE COLLECTED DATA ANALYSIS AND FUSION

Semantic Queries

A semantic annotation format facilitates data aggregation, and enables semantic inference engines to semantically fuse the data and process information for further analysis. In contrast to database queries where terminological knowledge plays no part in the query, our DL based representation enables the combined query of both terminological and assertional knowledge for greater situational understanding. Because data is captured about each assessment in the form of semantic annotations, those annotations can be leveraged to delve into multiple <https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2095-4>

dimensions of the data to examine particular subsets of the assessments.

For example, we can search by:

- `[[Category:Medical Survey]]` which shows all pages in a category such as Medical Survey.
- `[[Has disease::Cholera]]` which gives all pages annotated with the ‘Has disease’ property that has a property value of Cholera.
- `[[City::Saint Marc]]` which gives all pages annotated as being about the city Saint Marc.

Combining these annotations into one search would enable a user to identify Medical Surveys that have taken place in Saint Marc, where the individual interviewed was diagnosed with Cholera. Once we had this subset of data of interest, we could cross analyze the various dimensions of the data asked in the survey such as, where did the individual access their water. In this way, we are able to analyze across semantic dimensions within our data.

Semantic queries have several different options for result formats, such as maps, charts, graphs, timelines, calendars, and tables. Semantic queries are embedded within the pages of the semantic wiki, so that they are dynamically generated automatically with the most relevant and recent data incoming from the mobile collected forms, without requiring the user to actually understand and populate the semantic query syntax.

In the below example, we can see the result of a query about data collected about non-governmental organizations (NGOs). In this example, we query all NGOs in the knowledge base that are located in Thailand, viewing the associated URL for the organization, with the results displayed in a table format.

	URL
AACCU - Association of Asian Confederation of Credit Union	http://www.aaccu.coop/
AAT - Alliance Anti Traffic	http://aatthai.org/
ABCA - Asian Business Coalition on AIDS	http://www.allianceantitrafic.org/
ABV - Australian Business Volunteers	http://www.abconids.org/
ACET International - AIDS Care Education and Training Ltd	http://www.abv.org.au/
ACFOD - Asian Cultural Forum on Development	http://www.acet-international.org/
ACHR - Asian Coalition for Housing Rights	
ActionAid International	http://www.achr.net/
ACT - Organic Agriculture Certification Thailand	http://www.actionaid.org/
ADF - Asian Disaster Foundation	http://eng.actorganic-cert.or.th/
ADPC - Asian Disaster Preparedness Center	http://www.adf-thailand.org/
ADRA - Adventist Development and Relief Agency International	http://www.adpc.net/
ADRA (Thailand Asia Regional Office) - Adventist Development and Relief Agency International	http://www.adra.org/
AFC - Association of Finance Companies	http://www.adraasia.org/
AFD - Agence Française de Développement	http://www.afc.or.th/
AFECT - Association for Akha Education and Culture in Thailand	
AFET - Agricultural Futures Exchange of Thailand	http://www.afkathai.org/
AFI - Alliance for Financial Inclusion	http://www.afet.or.th/
AFIC - Asian Food Information Centre	http://www.afi-global.org/
AFPPD - Asian Forum of Parliamentarians on Population and Development	http://www.afic.org/
	http://www.afppd.org/

Figure 5. Dynamically generated table of NGOs and the property URL.

In constructing more complex visualizations, we encounter the need to conduct searches for properties that are semantic queries in and of themselves. In such visualizations, the information we need to query is not tagged within the pages, but by nesting a semantic query as a property value we can infer knowledge from the other semantic relationships that exist. For example, say there is a page for “Cholera” where we define a property whose value is assigned by a semantic query that returns a total for the number of cholera diagnoses aggregated from several locations. Because the property value is defined by a semantic query, it is a dynamic value. Thus, we can create an additional page within the semantic wiki to report an aggregation of various types of disease data for reporting, where the size of the bar on the pie chart may increase or decrease depending on the number of cholera diagnoses reported to the wiki at any given time, and where we can break out the analysis by the age of the patient.

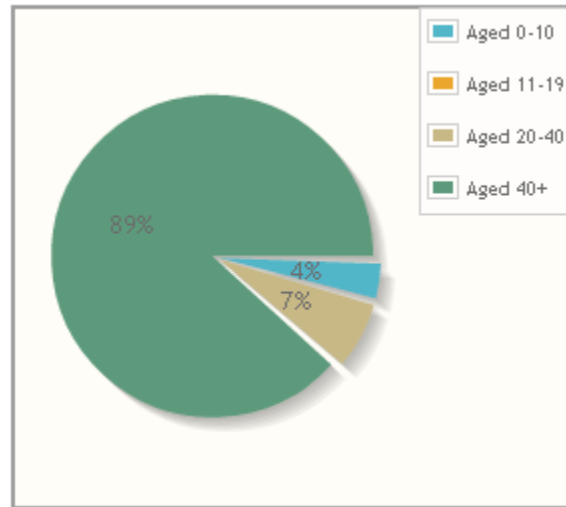


Figure 6. Dynamically generated pie chart of cholera diagnoses aggregated from a medical operation and broken out by patient age.

This is a query within a query, as we are delving into the multidimensional semantic relationships that exist, rather than the tags within a page, to deliver information. This innovative feature is a force multiplier as it demonstrates that visualizations and reports can be enhanced to drill down into multiple dimensions of data, querying for relationships nested among other relationships, to derive insight and produce refined visualizations in support of situational awareness.

Semantic Templates

By leveraging parser functions alongside semantic queries, we can also conduct statistical analyses. For example, we can create a template that embeds hundreds of individual semantic queries within the page that are all dynamically updated with the incoming mobile collected tagged data for data aggregation and analysis requirements. For example, perhaps we are interested in aggregating patient registration data for a medical support operation. These tables provide medical planners with on the ground analysis of how many patients have passed through medical registration throughout the day, the demographic information of these patients, as well as the most common treatments sought. Medical planners can then leverage this real-time information to plan medical relief accordingly.

As an example, we can create a template that allows a user to create a one-line table that depicts the most recent assessment submitted in any given category. To embed this template within a page, we can include the name of the template 'Most Recent Assessment' within double braces. When using the template, we include the name of the category of interest to view the most recent submission.

```
{{Most Recent Assessment | Category of Assessment}}
```

For instance, perhaps we are interested in viewing the most recent submitted assessment for a school assessment form.

```
{{Most Recent Assessment | School Assessment}}
```

A user can list additional qualifiers; for example, perhaps a user wants to view the most recent submitted school assessment, but only for those submitted for a particular location.

```
{{Most Recent Assessment | School Assessment | Location::Port-au-Prince}}}
```

For a more advanced analysis, templates with parameters produce different contents or have different behaviors depending on specifications provided by the user. For example, we can utilize a template to allow users to embed charts containing statistics about patient registrations completed by a particular team in a specific location as part of a medical assistance operation. Here we can see a chart that provides a patient breakdown by age.

Age	Total Patient Registrations	% of Patient Registrations
0-10 years	36	21.6%
11-19 years	7	4.2%
20-40 years	7	4.2%
41+ years	117	70.1%

Figure 7. Dynamically generated analysis of medical data from an operation using a template embedded on a page for automatic population of data as it is ingested from mobile device forms.

This template contains wiki syntax that constructs a formatted table with twelve semantic queries, however, a user would just need to transclude the template on a page rather than having to format the table and define each semantic query. To embed this template within a page, we include the name of the template we created such as ‘Medical Summary’ within curly double braces. When using this template, a user must pass parameters for location and medical team so that knowledge base understands which medical team and location the user is interested in.

```
{{Medical Summary|medical team=Team 1|location=Port-au-Prince}}
```

Geospatial Analysis

We utilize an interactive map for our geospatial analysis functionality to construct geospatial semantic queries. The user can write a query and see the results on a map, then refine the query and see the new results immediately, with no page refresh required. For example, we have defined the geospatial property “Within”, to translate to a semantic query, which provides all of the results whose coordinates fall within the polygon. The user does not have to understand this syntax however; they are able to select “draw polygon” from the map in the semantic wiki and draw a bounding area. From there the query is automatically generated representing the “Within” semantics. For example, here we can see the resulting forms that have been submitted from within the defined circle. Other geospatial operators that we have incorporated into our semantic search interface include “Contains”, “Covers”, “CoveredBy”, “Overlaps”, “Crosses”, “Disjoins”, “Equals”, “Intersects”, and “Touches”. This automated knowledge growth empowers users to monitor team activities, visualize impact, and geospatially analyze operations.

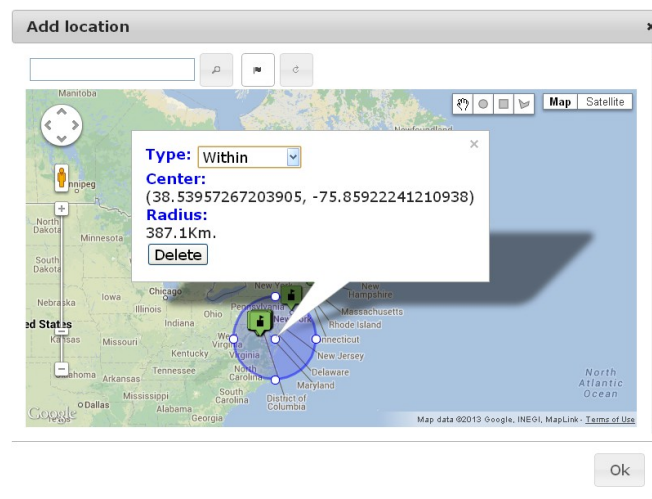


Figure 8. Semantic query incorporating geospatial semantics using the “within” operator.

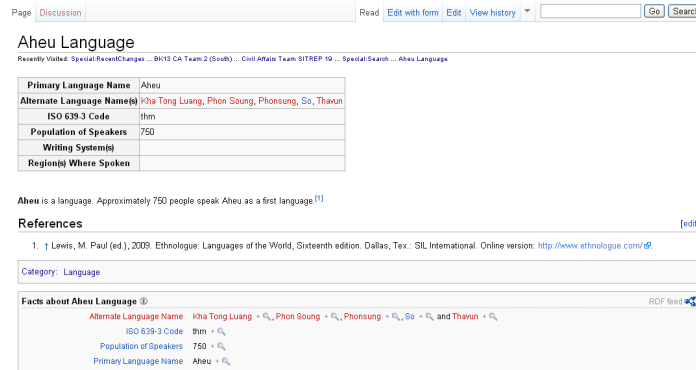
Linked Data

The linked data within the semantic wiki also allows users to browse associations among the mobile collected data. For example, the datatype listed in the XForm can specify that a form entry ought to be a page within the semantic wiki. That field will therefore contain an automatically generated link that directs the user to a page for that entity

<https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2095-4>

Cross-Cultural Decision Making (2019)

containing existing contextual knowledge. We have leveraged a number of contextual knowledge sources for thousands of pages in the wiki, automatically ingesting this information and producing the relevant annotations. For example, we semi-automatically ingested data pertaining to administrative divisions for particular countries, and annotated information within these pages about key leaders and social network data; cultural knowledge; and infrastructure information, and linked these locations to surveys completed by mobile data collectors. In this way, a user can submit an assessment in a particular area and then be able to go to that submission page on the semantic wiki and see a link to a page about that location with pertinent socio-cultural data. Here we can see a page referencing socio-cultural data that was automatically ingested into the semantic wiki.



Primary Language Name	Aheu
Alternate Language Name(s)	Kha Tong Luang, Phon Soung, Phonsung, So, Thavun
ISO 639-3 Code	thm
Population of Speakers	750
Writing System(s)	
Region(s) Where Spoken	

Aheu is a language. Approximately 750 people speak Aheu as a first language^[1]

References [edit]

- Lewis, M. Paul (ed.), 2009. Ethnologue: Languages of the World, Sixteenth edition. Dallas, Tex.: SIL International. Online version: <http://www.ethnologue.com/>

Category: Language

Facts about Aheu Language RDF feed

Alternate Language Name [Kha Tong Luang](#) [Phon Soung](#) [Phonsung](#) [So](#) [Thavun](#)

ISO 639-3 Code [thm](#)

Population of Speakers [750](#)

Primary Language Name [Aheu](#)

Figure 9. Aheu Language page automatically ingested from Ethnologue Website (<https://www.ethnologue.com>) for linking socio-cultural information to mobile collected data.

It is important to note that when items are blue within the semantic wiki, it means that there is a page about that entity and that there is content populating that page. When items are red, it denotes that the wiki has indicated that this entity is a page, but there is no content on that page. This type of demarcation can assist operators with data collection requirements. For example, if a user sees that a particular entity is referenced numerous times in the wiki, but that there is no content supporting the understanding of this entity in the knowledge base, meaning it is marked as red, the user can prioritize data collection requirements against this entity. The semantic wiki provides a list of the most referenced red entities to assist in prioritizing data collection requirements.

Developing methods and means for automatically ingesting relevant socio-cultural data from websites and databases, tagging the information, and representing it semantically within our semantic wiki was one of the most significant developments for our solution as it enables the augmentation of mobile collected data by automatically linking to relevant socio-cultural data for situational and threat awareness.

CONCLUSIONS

Our semantic knowledge management solution for the semantic enrichment and fusion of mobile collected data, *SemLayer*, eliminates the challenge of form management for mobile data collection campaigns, providing a cloud service for collection, aggregation, analysis, and collaboration in support of big data analysis. We use semantic annotations as a semantic enrichment representation for knowledge management, leveraging machine automation to capture semantic enrichment through form based data collection and user defined semantic annotations in the wiki text. Semantic forms in the wiki capture semantic annotations for semantic enrichment allowing the knowledge base to automatically link categories, properties, and values in the data collected with contextual knowledge captured in the knowledge base ontology ingested from other data repositories. Semantic inferencing can be effectively used to infer semantic enrichments for greater data aggregation and analysis.

What is new about our approach is that we enable authoring of form semantics; automate form publishing and distribution using semantic technology; and automate semantic enrichment and the semantic fusion of mobile collected data. Semantic technology enables the use of spreadsheets to define the semantics of a form, and upload it to the knowledge base for automated generation of semantic forms. Upon upload, the form is automatically distributed to mobile devices while any changes to the forms are automatically propagated. Form submissions are

automatically ingested into the knowledge base, and linked with contextual knowledge. The semantic representation enables the use of semantic queries to analyze and gain insight in the collected data, while spatial semantics are fully integrated with semantic queries, which can be specified and visualized using maps. Semantic queries fuse data across different data sources, across different data dimension types, and across different dimension instances.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research under the direction of Dr. Ivy Estabrooke and Gary Kollmorgen.

REFERENCES

- Caglayan, A. (2013, June). "Mobile Enabled Spatially Aware Semantic Wiki for Information Management using Open-Source Software Platforms", Presentation at Semantic Technology and Business Conference, San Francisco, CA.
- Caglayan, A., Cassani, L., Mooney, L., Morgan, W., Boyles, A. (2012, November). Semantic Technologies for Civil Information Management during Complex Emergencies. In *2012 IEEE Conference on Technologies for Homeland Security (HST)* (pp.523-528). Boston, MA.
- Das, S. (2009), *High-Level Data Fusion* Norwood, MA: Artech House.
- El Ghali, A., Tifous, A., Buffa, M., Giboin, A., Dieng-Kuntz, R. (2007, September). Using a Semantic Wiki in Communities of Practice. In *Proceedings of the 2nd International Workshop on Building Technology Enhanced Learning Solutions for Communities of Practice* (pp. 22-31), Crete, Greece.
- Horridge, M., Bauer, J., Parsia, B., Sattler, U. (2008, October). Understanding Entailments in OWL. In *Proceedings of the Fifth OWLED Workshop on OWL: Experiences and Directions*. Karlsruhe, Germany.
- Melanson, D. (2011), *UN: worldwide internet users hit two billion, cellphone subscriptions top five billion*. Engadget Website: <http://www.engadget.com/2011/01/28/un-worldwide-internet-users-hit-two-billion-cellphone-subscript/>
- Nardi, D., Brachman, R.J. (2003), "An introduction to description logics", in: *The Description Logic Handbook*, Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (Eds.). pp. 1-40
- Schaffert, S., Bischof, D., Bürger, T., Gruber, A., Hilzensauer, W. (2006, June). Learning with Semantic Wikis. In *Proceedings of the First Workshop on Semantic Wikis: From Wiki to Semantics (SemWiki2006)*. Budva, Montenegro.
- Stroh, L., Caglayan, A., Rashed, T., Burke, D., Eaton, G. (2010), "Geospatial Campaign Management for Complex Operations", in: *Advances in Cross-Cultural Decision Making*, Schmorow, D., Nicholson, D. (Eds.). pp. 460-469
- Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F. (2006). "Semantic annotation for knowledge management: Requirements and a survey of the state of the art". *Web Semantics Volume 4 No. 1*.