SMART AND CONNECTED HEALTH

Humanitarian Assistance Ontology for Emergency Disaster Response

Satria Hutomo Jihan and Aviv Segev, *Korea Advanced Institute of Science and Technology (KAIST)*

A method for merging ontologies and logic rules represents humanitarian needs and recommends appropriate responses during a crisis. This automatic processing of crisis data provides useful information to decision makers.

When a crisis arises, we must determine actual needs before taking action so as to quickly provide proper humanitarian assistance in such emergencies. However, decision makers face the challenge of identifying crisis needs from the massive information about the crisis that might appear at once and in an unstructured format. Unstructured information is unannotated data intended only for basic readability. The unstructured information from different sources might also contain different expressions for similar needs. These issues can hinder machines that automatically process and transform information into a useful knowledge representation that depicts the actual crisis needs.

After the 2011 earthquake and tsunami in Japan, local people used social media websites, such as Twitter and Facebook, to communicate with relatives and to keep in touch with the world, because the mobile phone networks in Japan were overloaded. A study conducted by the American Red Cross in 2009 found that social media websites are one of the most popular sources for access to emergency information.[1] This online data could create new opportunities for addressing emergency response.

Ontologies have been used to provide structured information that creates meaningful relationships between information resources and to allow machines to process, infer, or combine the information from different sources automatically into a consistent body of knowledge. In crisis response, ontologies can unify data from different resources syntactically and semantically. However, the related domain experts who are required to construct the ontology might not be available during the crisis.

This article addresses the critical issue that humanitarian response practitioners have to deal with regularly. While lacking relevant information, they get flooded by unstructured information that is difficult to process and analyze. Humanitarian response becomes even more important as new sources of unstructured data, such as social media, become available. We propose a method for automatic extraction of ontological information from unstructured crisis data.

The purpose of this study is to build a Humanitarian Assistance Ontology (HAO)

1541-1672/14/$31.00 © 2014 IEEE
Published by the IEEE Computer Society

Related Work

An ontology is defined as an explicit and formal specification of a conceptualization.[1] An ontology consists of a list of terms and their relations to represent the domain. Ontologies have been used to provide shared and common understanding of a domain and can facilitate interoperability among information systems and also share and reuse knowledge among systems.

Ontologies can be built either from scratch[2] or by reusing existing ontologies and updating them for the current need. Contexts can also be used to verify relationships among ontology concepts.[3] However, the method of building a proper ontology is still an open issue depending on its domain.

In crisis response, ontologies have been used in several ways. Paolo Di Maio[4] addresses the open ontology methodology for open source emergency response systems. An open ontology allows users and developers to collaboratively and dynamically create and support knowledge and semantic consistency for emergency response systems, as opposed to a closed ontology, whereby the ontology is developed by an organization to impose a single view of the world without a public consultation process and deliverable.

Xiang Li and his colleagues[5] propose a practical emergency response workflow and emergency response ontology architecture. In their paper, the proposed ontology's goals are to standardize semantic concepts that can be applied to many different emergency response systems and to define practical common vocabularies between emergency response personnel.

Zhengjie Fan and Sisi Zlatanova[6] explore the semantic interoperability of the terms and spatial information to be used by different emergency response communities.

These works didn't aim to create a solution based on humanitarian standards to create a bridge between the emergency needs and the crisis response recommendations using an ontology-merging process to represent the knowledge and rule-based techniques to support the decision-making process.

References

1. T. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," *Formal Ontology in Conceptual Analysis and Knowledge Representation*, vol. 43, nos. 5–6, 1995, pp. 907–928.
2. F. Belleau et al., "Bio2RDF: Towards a Mashup to Build Bioinformatics Knowledge Systems," *J. Biomedical Informatics*, vol. 41, no. 5, 2008, pp. 706–716.
3. S.H. Jihanand and A. Segev, "Context Ontology for Humanitarian Assistance in Crisis Response," *Proc. Int'l Conf. Information Systems for Crisis Response and Management* (ISCRAM-13), 2013.
4. P. Di Maio, "An Open Ontology for Open Source Emergency Response System," *Open Source Research Community*, Jan. 2007; www.academia.edu/610005/An_Open_Ontology_for_Open_Source_Emergency_Response_System.
5. X. Li et al., "Building a Practical Ontology for Emergency Response Systems," *Proc. Int'l Conf. Computer Science and Software Engineering*, vol. 4, 2008, pp. 222–225.
6. Z. Fan and S. Zlatanova, "Exploring Ontologies for Semantic Interoperability of Data in Emergency Response," *Applied Geomatics*, vol. 3, no. 2, 2011, pp. 109–122.

for emergency disaster response based on the International Emergency Response standard by using online expert resources and blog media data. We introduce the standardized ontological model and methodology to use actual crisis blog data from the Web.

The main contributions of this work, compared to previous works, are that it:

- Describes a method for construction and instantiation of generic crisis identification and response ontology using expert resources.
- Presents methods for ontological crisis concepts extraction from online textual data.
- Constructs a technique for inference logics that might derive recommendations based on humanitarian assistance standards.
- Implements a knowledge representation system for rapid emergency response.

For others' work in this area, see the related sidebar.

Method: Humanitarian Assistance Ontology (HAO)

To harness the massive amounts of information from humanitarian open data, input data needs to be transformed into useful information for the decision makers, and the system needs to be updated to reflect the current situation. The manual processing of all information is impractical, because considerable time and manual resources are required. A crisis response system should process the information and assist in providing rapid response in a short time period. In particular, when a disaster occurs, the system automatically captures related information and relevant context and delivers its recommended responses from which decision makers can choose.

The proposed HAO method described in Figure 1a consists of a preprocessing stage and two main inference parts: the Crisis Identification Ontology and the Crisis Response Ontology. The Crisis Identification Ontology is composed of two parts: the global part containing predefined static concepts, and the local part containing dynamic concepts that can be continuously updated according to current situations (see Figure 1b). For instance, the global ontology might include the class "crisis needs," which covers all possible humanitarian needs that can be used as the range domain of property "impact." The local part, conversely, includes

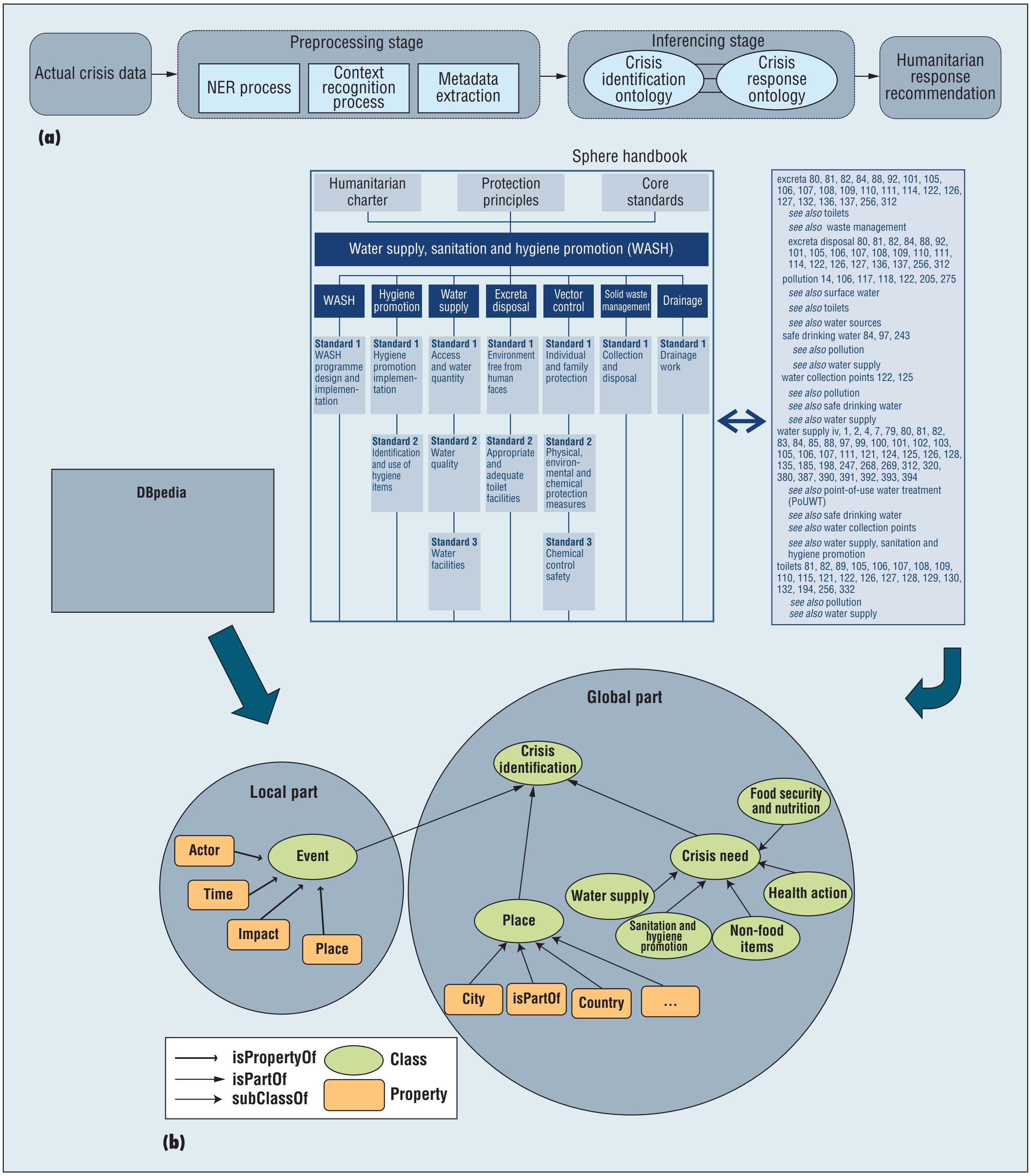


Figure 1. Overview of a crisis response system. (a) Crisis identification method and (b) the Crisis Identification Ontology.

information about specific events that occurred in a certain place with their crisis impacts.

Crisis Identification Ontology

The main purpose of the Crisis Identification Ontology is to assist in the categorization of the crisis information based on the existing domains. The global part of the ontology can be applied regardless of the crisis type. While the global part of the ontology remains static during the event, the content of the local part can be changed continuously during the flow of crisis information. The benefits of using the global part of the ontology are to support ontology reuse and to reduce ontology construction time.

Figure 1b exhibits several classes in the Crisis Identification Ontology. The classes "crisis need" and "place" are categorized as in the global part, because they can be constructed using a generic data source such as DBpedia and *The Sphere Handbook*. DBpedia is the community-driven data pool that structures data from Wikipedia. DBpedia relies on the existing Wikipedia community to update its content. To keep its knowledge base up-to-date according to Wikipedia, DBpedia uses extraction frameworks.[2] Some applications are publicly available using DBpedia identifiers, such as Zemanta (see www.zemanta.com) and Faviki (see http://faviki.com). *The Sphere Handbook* (see www.sphereproject.org/handbook) is one of the most widely known and internationally recognized sets of common principles and universal minimum standards in humanitarian response areas. *The Sphere Handbook* was designed to improve a humanitarian agency's quality of humanitarian assistance and its accountability to constituents, donors, and affected populations. It should be noted that the global part of the Crisis Identification Ontology is a set of predefined concepts constructed before the crisis, and it can be expanded by adding other disaster recovery resources, such as the *Operational Guidelines and Field Manual on Human Rights Protection in Situations of Natural Disaster* (see www.refworld.org/docid/49a2b8f72.html).

We adapted relations of the class "crisis need" and its subclasses from *The Sphere Handbook*'s table of contents. We used DBpedia to derive several properties, such as city, country, and so on, in the class "place." Each class of the global part of the Crisis Identification Ontology is used as data references. For example, the class "crisis need" and its subclasses are used as a reference to identify humanitarian needs of extracted terms. Another example is the property "isPartOf," which is used to expand the place of an event into a greater scope of area.

The local part contains the dynamic data based on the actual crisis data flow. We define the class "event" to capture the current crisis situation. The "event" class includes at least four properties:

- actor—the person reporting about the crisis occurrence place and involved with the crisis event;
- time—the time indicating when the report was submitted;
- place—the location indicating where the crisis impact occurred; and
- impact—the string description about the main message from the event.

When a textual document containing crisis information enters the HAO system, the system preprocesses it by extracting required information and transforming it into one instance of the class "event." To acquire all required information needed for properties of the class "event," we apply several approaches, such as the name-entity recognition (NER) process for extracting the document's place and actor information, the context recognition algorithm for eliciting impact information, and the document metadata extraction for time information describing the document's creation time.

Crisis Response Ontology

To make the initial ontology, indexes from *The Sphere Handbook* and the chapter structures can be extracted to build hierarchical and semantic relations (see Figure 1b, top right). Each index can be an ontology class. For example, the index "water supply" can become a class, which has relations with the classes "PoUWT" (or point-of-use water treatment), "safe drinking water," "water collection points," and "toilet."

Each of these Crisis Response Ontology classes has class properties consisting of *The Sphere Handbook* page number and the predefined values from the page. For example, the class "water supply" has many related pages that describe in detail requirements of humanitarian assistance such as the requirement of average water use for drinking, cooking, and personal hygiene, and maximum distance from any household to the nearest water point.

Ontological Crisis Concepts Extraction Methods

The HAO system takes actual crisis data as the initial input. Actual crisis data consists of textual documents that can be obtained from blogs, emails, or other open data on the Internet. Each document is regarded as one dataset that can be preprocessed for the system input.

The preprocessing of actual crisis data into input data for the system requires three kinds of processes: NER, context recognition, and metadata extraction.

The NER process refers to the extraction of words and strings of text within documents that represent discrete concepts, such as names and places. We can use various tools and methods to perform NER processes; we use the Stanford NER here because it has fairly robust results and

```
r1: EVENT(?ev), hasPlace(?ev, ?p), hasImpact(?ev, ?imp)→RECOMMENDATION(?p)
r2: EVENT(?ev), hasImpact(?ev,?imp), hasPlace(?ev,?p)→hasImpact(?p, ?imp)
r3: hasImpact(?p, ?imp)→needs (?p, ?imp)
r4: EVENT(?ev), hasImpact(?ev, ?imp), hasPlace(?ev, ?p), responseType(?imp,?rt)→response(?p,?rt)
r5: hasReference(?imp,?sp), responseType(?sp, ?x)→responseType(?imp, ?x)
r6: EVENT(?ev), hasImpact(?ev,?imp), seeAlso(?imp,?imp2), responseType(?imp2,?rt2)→response(?p,?rt2)
r7: EVENT(?ev), hasPlace(?ev,?p), populationTotal(?p,?pt), hasImpact(?p,?imp), hasReference(?imp,?sp),
    basicWaterNeed(?sp,?bws), multiply(?y,?pt,?bws)→totalWaterNeed(?p,?y)
```

Figure 2. Inference rules of the Humanitarian Assistance Ontology (HAO).

outperforms OpenNLP (NLP stands for natural language processing).[3] Through the NER process, we can obtain the required place information from a document.

The context recognition process refers to identifying the important contexts of documents from an actual crisis datum relating to the humanitarian need. As a first stage of context recognition, the actual data is processed into a set of terms using the C-value algorithm.[4]

After we obtain a set of terms using the C-value algorithm, the contexts of each term can be recognized by using the context recognition algorithm. In this algorithm, the contexts of each term are obtained using the Internet—which acts as a context database—and then ranked. Each term is regarded as a query to an Internet search engine. The preliminary contexts of each term are obtained by implementing the term frequency method on the webpages retrieved.

To determine humanitarian needs from a document, the extracted terms of that document need to be mapped into the global part of the Crisis Identification Ontology. Let d_i be the extracted terms of a document and $D = \{d_1, d_2, \ldots, d_n\}$ be a set of textual descriptors representing documents where each d_i has a set of contexts $X_i = \{x_1, x_2, \ldots, x_p\}$ describing all possible scenarios. To map the extracted terms into the Crisis Identification Ontology, the string matching function between the terms' contexts and ontology's concepts is used. The function is denoted by $\text{match}_{\text{str}}$, which returns 1 if two strings match and 0 otherwise. Let $O = <C, R>$ be a simplified representation of an ontology where $C = \{c_1, c_2, \ldots, c_n\}$ is a set of concepts with their associated relation R. The match between terms' context and ontology is defined as follows:

$$\text{match}(d_i, O) = \sum_{x_p \in d_i} \sum_{d_i \in D} \sum_{c_n \in C} \text{match}(x_p, c_n).$$

If the contexts are matched with any classes in the global part of the Crisis Identification Ontology, those classes are recognized as required humanitarian needs representing that document. The results of the combination of the result of the context recognition process (the main message of humanitarian needs of a document) with the result of the NER process (actor and place information from a document) and with metadata of the time when that document is created become the properties of an instance of the class "event," which describes a crisis event in a specific place with specific humanitarian needs.

By using a set of predefined logic rules, the HAO system processes instances from the class "event" and infers corresponding responses from the Crisis Response Ontology. The logic rules mechanism has three functions: to enable the system to automatically derive recommendation statements from a set of premises/facts, to provide explanations of how recommendations are made, and to create a bridge between the Crisis Identification Ontology and the Crisis Response Ontology. Using this information, decision-makers are able to prioritize actions according to their perspectives and limitations. Figure 2 shows the general inference logic rules of HAO.

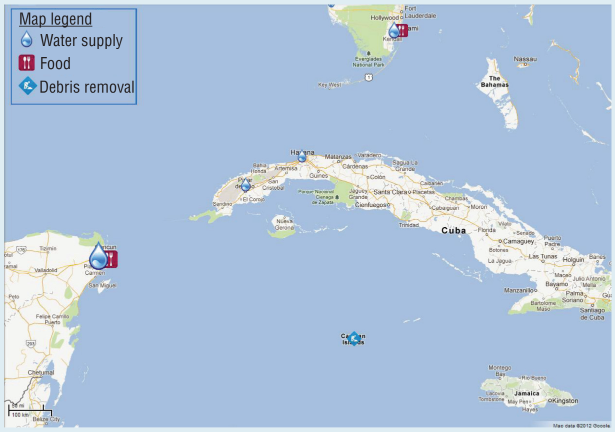
Implementing HAO: The Example of Hurricane Wilma

We chose Hurricane Wilma as the case study for this article, because it was a multinational crisis in which a massive amount of information came out in a short period of time. Hurricane Wilma was one of the most intense hurricanes to occur in the Atlantic Basin. Hurricane Wilma formed on 15 October 2005 and dissipated on 26 October 2005. It affected a wide area, including Cuba, Jamaica, Honduras, Bahamas, and Florida.

The actual crisis data of Hurricane Wilma came from blogs, community sites, or news websites. The following report is taken from bbc.co.uk and describes actual crisis conditions of people at that time:

My sister and her husband are amongst 1,000 British tourists that have been abandoned by their travel companies, and are currently in a car park somewhere in Cancun. These people have no running water, no toilets, no food, no clean clothes, and have no idea what is happening to them, and when they are going to be able to come home. What is being done to help them?

Adam Norris, Chichester, 28 October 2005



No	Place	The Humanitarian Needs and Priorities
1	Pinar del Rio	Water Supply(1), Drainage(1), Excreta Disposal(1), Settlement(1)
2	Havana	Water Supply(1), Drainage(1), Excreta Disposal(1), Settlement(1)
3	Cuba	Debris Removal(1)
4	Florida	Water Supply(1), Debris Removal(2), Food(2)
5	Cancun	Shelter(1), Water(1), Food(2), Clothing(3), Toilet(3), Drainage(4)
6	Isla Mujeres	Water Supply(1), Clothing(2)
7	Jamaica	Debris Removal(1), Flood Recovery(1), Shelter(1)
8	Miami	Water Supply(1), Debris Removal(2), Food(2)

Figure 3. Distribution map of humanitarian needs and priorities. The larger icons indicate a greater priority.

After the data preprocessing by the NER process and context recognition system, an instance of the class "event" is created. To address event instances, the system infers the possible actions from the Crisis Response Ontology. Rule r1 infers Cancun as a place instance in the recommendation class, because that place was affected by crisis. Rule r2 aggregates the crisis impacts on Cancun from many other reports, and rule r3 summarizes response needs based on identified crisis impact. Rules r4 and r5 generate the possible humanitarian responses; in this case the recommendations of humanitarian action are emphasized on food security and nutrition, water supply, non-food items, and excreta disposal. Rule r6 infers total water needs for the Cancun area (9,424,590 liters per day) by using the amount of basic survival water needs in water supply response (15 liters per day) and the DBpedia property such as areaTotal and populationTotal of Cancun. Last, by going to the sections of *The Sphere Handbook* indicated by HAO, decision makers can obtain an explanation of the recommendations suggested by the system. This aggregation can help decision makers send humanitarian needs in bulk effectively and efficiently. In addition, we can obtain an overview of humanitarian needs distribution to evaluate the effect of that crisis.

Experiments

Using our approach, we analyzed a dataset of 125 textual reports (each with an average of 114 words) from social media blogs in bbc.co.uk and cnn.com where people voluntarily gave information about what was happening during the Hurricane Wilma crisis. These randomly collected reports were posted by people who were affected by or had relatives in the affected area to update the current status of crisis as it occurred. The ontology for the experiments included 40 concepts and 42 relations. The report datasets mentioned 107 locations. We selected 10 locations for precision and recall analysis based on the number of reports, because these locations had three or more crisis needs.

If duplicate data comes from the same person and reports the same crisis need for the same place, we can use a simple algorithm to check and eliminate duplicate reports. If the data comes from different people but reports the same crisis need for the same place, the need isn't eliminated and receives a higher priority because the crisis need is more prominent.

If the data from social media provides a field report, which is a live report containing more detailed information of an occasion during a crisis from each person who contributed, the data from the news media often provides a summary or the general information of the crisis event. When decision makers receive news data, such as the number of people migrating during a crisis, from other sources, they can tune the Crisis Identification Ontology by creating an instance containing the updated number of affected people in the class "place." Later, the system can use this updated information to infer new possible actions for decision makers.

Using Google Maps, the places having specific icons, such as for food, water supply, or debris removal, indicating needs for humanitarian assistance for those places. The bigger the icons, the greater the priority. Figure 3 exhibits the humanitarian response needs for the specific areas of Cancun, Cuba, Cayman Islands, and Florida.

The table in Figure 3 displays more examples of humanitarian needs in various places during Hurricane Wilma. The numbers beside the

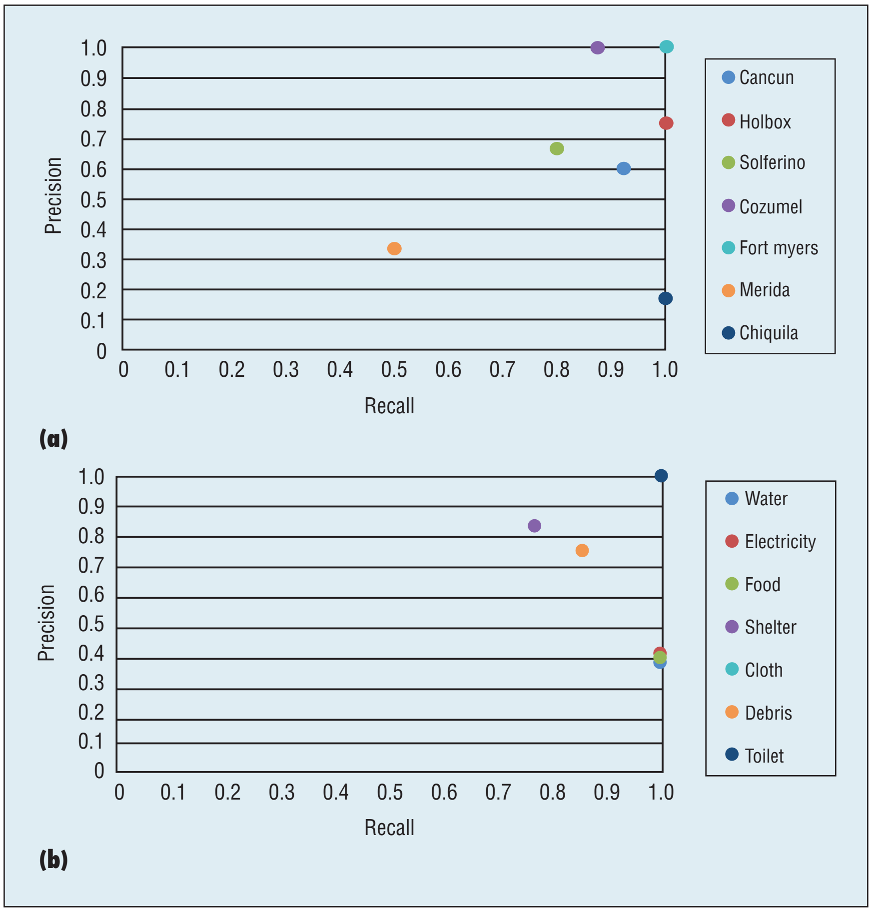


Figure 4. Precision and recall evaluation per (a) location entity and (b) crisis need entity. In (a), the system has a relatively high recall score for every location except Merida. In (b), cloth and toilet have the same precision and recall scores; thus their dots are overlapping. Water, electricity, and food entities have the same recall score and close precision scores, and thus their dots intersect.

Table 1. Overall performance comparison between methods.

	Evaluation				
Methods	Recall	Precision	F-score	Location F-score	Need F-score
Humanitarian Assistance Ontology	0.897	0.530	0.666	0.493	0.753
Keywords matching	0.744	0.483	0.585	0.362	0.681
AlchemyAPI	0.051	0.333	0.089	0.083	0.073

type of needs indicate the priority of humanitarian response. Humanitarian need with number (1) means it has highest priority. These priorities come from the weights of each possible action needed to assist the corresponding humanitarian needs.

We evaluated the system result by comparing it with datasets of all actual crisis needs in every location collected by two graduate students. If the result of the system matched the result of at least one evaluator, then we considered the result. Figure 4 shows the evaluation result for each location and each crisis need.

Figure 4a shows seven location entities. Three other locations (Mexico City, Miami, and New Orleans) are missing in the graph because they don't have actual crisis needs (false positive error). The HAO system has a relatively high recall score for every location except Merida. Figure 4b shows seven crisis need entities. Cloth and toilet have the same precision and recall scores; thus their dots are overlapping. Water, electricity, and food entities have the same recall score and close precision scores, and thus their dots intersect in the graph.

We defined the keywords-matching method and AlchemyAPI (see www.alchemyapi.com) as comparison methods to evaluate our proposed methods. The input to all three methods was the free text collected from the blogs. For keyword-matching queries, we based the datasets on predetermined crisis need types and location names to identify whether a particular location has the crisis need keywords. We calculated keyword matching based on whether crisis need and location keywords appear in each report. AlchemyAPI is a popular natural language processing service, providing users with content analysis and metadata annotation tools such as semantic metadata extraction about people, places, topics, and so on. For AlchemyAPI, the entities under Country, City, and StateOrCity tags become location entities, while the entities under Concept tags become potential crisis entities. Then location entities and potential crisis entities of each report are compared, and their appearance in each report increases the number of crisis need occurrences identified by AlchemyAPI.

Table 1 shows the performance scores of three methods of detecting crisis needs when the methods are implemented. The HAO method outperforms the other two methods in recall, precision, F-score, location F-score, and need F-score. All methods have relatively low precision scores, thus indicating the common challenge of a machine system to identify actual needs from the textual report. AlchemyAPI has the lowest recall score, because it only retrieved a few location tags with the crisis needs tags.

These low results could be caused by each method's features. The HAO method has a higher recall than keyword matching because it expands the terms to include their context so that the HAO method can detect more crisis needs. The keyword-matching method has a higher result than AlchemyAPI because the keyword-matching method uses a predetermined list of keywords to extract entities from datasets, whereas AlchemyAPI uses NER, statistical NLP, and machine-learning algorithms to extract the entities.

We qualitatively compared the recommendation offered by the HAO system to the real crisis situation to see whether the recommendations can fulfill actual needs. For the evaluation, we used the *Operations Update on Hurricane Wilma* (see www.ifrc.org/docs/appeals/05/05EA02401c.pdf) issued on 27 January 2006 by the International Federation of Red Cross and Red Crescent Societies.

Compared to the needs identified by the HAO system, three out of five identified needs for Cozumel were fulfilled: food, shelter, and electricity. The other two needs weren't mentioned in the report (water and debris removal). For Cancun, food and hygiene kits were delivered. However, compared to the needs identified by the HAO system, six other crisis needs (cloth, electricity, shelter, toilet, water, and debris removal) aren't mentioned in the actual humanitarian operational update. This discrepancy could imply that the relief assistance undertaken wasn't enough to meet actual humanitarian needs. The crisis needs identified by the HAO system could reveal unfulfilled humanitarian needs and act as a reference for humanitarian organizations providing relief assistance to meet minimum humanitarian standard needs.

The HAO system, as a decision support tool, utilizes existing domain expert resources, such as DBpedia and the Sphere Handbook, to process the actual crisis data and represent humanitarian needs in specific areas with recommended actions and their priorities in the response to the needs. By unifying logic rules with the Crisis Identification and Crisis Response ontologies and automatically generating recommendations, the system aids with the process of quick humanitarian response. Future directions of research include an expanded analysis of the number of locations, comparison of system performance for different crisis types, and connection of the HAO system for implementation purposes.

THE AUTHORS

Satria Hutomo Jihan is a technology consultant in Indonesia. His research interests include ontology, e-learning, and knowledge consulting. Jihan has an MS in knowledge service engineering from the Korea Advanced Institute of Science and Technology (KAIST). Contact him at satriahj@kaist.ac.kr.

Aviv Segev is an associate professor at the Knowledge Service Engineering Department at the Korea Advanced Institute of Science and Technology (KAIST). His research interests include knowledge classification using the Web; context recognition and ontologies; knowledge mapping; and implementations of these areas in the fields of Web services, medicine, and crisis management. Segev has a PhD in technology and information systems from Tel-Aviv University. He's a member of IEEE and the ACM. Contact him at aviv@kaist.edu.

Acknowledgment

This work was supported by the IT R&D program of MSIP/KEIT (grant no. 10044494, WiseKB: Big data-based self-evolving knowledge base and reasoning platform).

References

1. B. Lindsay, "Social Media and Disasters: Current Uses, Future Options, and Policy Considerations," *Congressional Research Service*, report R41987, 2011; www.infopuntveiligheid.nl/Infopunt-documenten/R41987.pdf.
2. C. Bizer et al., "DBpedia—A Crystallization Point for the Web of Data," *Web Semantics Science Services and Agents on the World Wide Web*, vol. 7, no. 3, 2009, pp. 154–165.
3. F. Nuno, J. Borbinha, and P. Calado, "An Approach for Named Entity Recognition in Poorly Structured Data," *The Semantic Web: Research and Applications*, LNCS 7295, Springer, 2012, pp. 718–732.
4. K Frantzi, S. Ananiadou, and H. Mima, "Automatic Recognition of Multi-Word Terms: The C-Value/NC-Value Method," *Int'l J. Digital Libraries*, vol. 3, no. 2, 2000, pp. 115–130.