A Database Driven Initial Ontology for Crisis, Tragedy, and Recovery

ABSTRACT

Many databases and supporting software have been developed to track the occurrences of natural disasters, man-made disasters, and combinations of the two. Each of the databases developed in this context, define their own representations of a disaster that describe the nature of the disaster and the data elements to be tracked for each type of disaster. The elements selected are not the same for the different databases, yet they are substantively similar. One capability common to many ontology development efforts is to describe data from diverse sources. Thus, we began our ontology development process by identifying several existing databases currently tracking disasters and derived the "ontology in situ" of their database. That is, we identified how the designers of the databases classify the types of disasters in their systems. We then merged these individual ontologies to identify an ontology that includes all of the classifications from the databases. Several aspects of disasters from the databases were highly consistent and therefore fit well together, e.g., the types of natural disasters, while others, e.g., geographic descriptions, were idiosyncratic and do not fit together seamlessly. The resulting ontology consists of 185 elements and has the potential to support data sharing/aggregation across the databases considered.

Keywords

Ontology development, crisis, database schema.

INTRODUCTION

The Crisis, Tragedy, and Recovery Network (CTRnet) project has been collecting news and online resources that are related to natural disasters (e.g., wildfires, floods, typhoons, and earthquakes) and man-made tragedies (e.g., campus shootings in the USA and internationally). Since the summer of 2010, the project group also has been working on collecting and providing CTR-related information from social media such as Twitter and Facebook. Our goal for the development of the CTRnet digital library (DL) is to collect, organize and serve resources that can cover the disaster and emergency management domain comprehensively. Without knowing which concepts are important and how they are related with one another, we may not see the big picture. This might lead the CTRnet DL to collect and provide resources that are unbalanced in covering the domain. For this reason, we need are developing a CTR ontology with in-depth coverage [12].

The CTR domain is broad. Therefore, having solid domain knowledge will help to collect and organize resources. It also will help visitors navigating through information in the CTRnet digital library [4]. Developing a CTR ontology will allow us to assemble a key form of domain knowledge. To confirm accuracy across the broad CTR area, and to make the effort feasible and scalable, it makes sense to create the ontology through a semi-automatic methodology involving human as well as computational efforts (e.g., natural language processing).

LITERATURE REVIEW

ISCRAM has often included papers addressing the use of ontology in the disaster context. The CTR ontology will cover the domain of both natural and man-made disasters and emergencies. We plan to use it in two ways. One is to use the concepts in the ontology as a taxonomy list. The disaster resources in the CTRnet digital library can be tagged in more formal and comprehensive way instead of being tagged with rather ad-hoc user-generated tags. This will promote the increased shared understanding of the resource description among the digital library patrons. The other is to use the ontology in the resource classification. An automated process that takes information gathered through web-crawling and assigns to that information tags from the ontology.

RESEARCH METHOD

We have bootstrapped the ontology development process by merging the classification variables from four databases designed to record the existence of disasters. We selected this method due to the commitment in resources required to develop a database and supporting information system. We assume those investing in the design and development of such systems would uncover the essential elements of disasters in their requirements analyses and design their systems to support those elements. By evaluating multiple databases we hope to identify elements that are common as well as elements more idiosyncratic to particular systems. The databases we evaluated included the Richmond Disaster database (http://learning.richmond.edu/disaster/index.cfm), the EM-DAT database (http://www.emdat.be/), the Canadian disaster database http://www.publicsafety.gc.ca/res/em/cdd/search-en.asp), and the DesInventar tool for defining disaster databases (http://www.desinventar.org/). The resulting draft ontology current consists of 185 elements. Our process for merging the disaster classifications from the four databases was straightforward.

We started with the disaster classification form in the Richmond database, then merged into it the EM-Dat database classification elements. The process went smoothly with the Richmond database providing more leaves and the EM-Dat elements comprising more high level elements. Next we merged the Canadian disaster database into the ontology, assigning each of its elements to an existing element in the draft ontology or creating a new element in the ontology. The overall hierarchy of the draft ontology was stable through this process with most of the Canadian disaster database mapping to leaf elements in the ontology. Finally we merged the DesInventar disaster classification into the merged ontology from the other three databases. A large majority of elements matched the draft ontology. However, the concept of the "cause" of the disaster was not included in any of the other databases. At this point we have decided to exclude this element from the merged ontology due to the lack of consensus across databases. However, DesInventar includes an extensive set of slots to be filled for every disaster and we have adopted them for integration with the ontology.

RESULTS

Figure 1 shows the highest level of the ontology. There is substantive consensus across all the databases we evaluated for a "manmade" versus "natural" type of disaster at the highest level of describing disasters. Figure 2 shows how the ontology expands as elements at higher and mid levels are selected. In this example Natural disasters are expanded to show those that are Biological, Climatological, and Geophysical. Meterological disasters are floods including Flash floods, general flooding, and coastal flooding, and are not shown in Figure 2.

Similar figures can be made for the Manmade types of disasters, but are not presented for space considerations. Manmade disaster elements are 140 of the 185 elements in the ontology, perhaps suggesting a focus of those trying to track disasters.
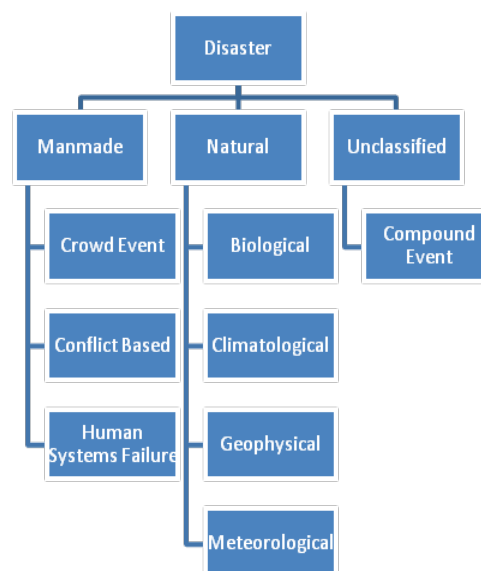
Figure 1: Highest level of ontology merged from four disaster database projects.

       Another characteristic of the ontology shows the bias of the users of the disaster databases toward the need to describe some elements in greater detail than others. For example, a large variety of storm conditions were identified. Figure 3 shows the fifteen types of storms listed, which are substantially greater than the number of sub-elements for most other elements, demonstrating a finer granularity of attention to this topic versus others. Similarly, Terrorism also has substantially more elements than most others as shown in Figure 4.
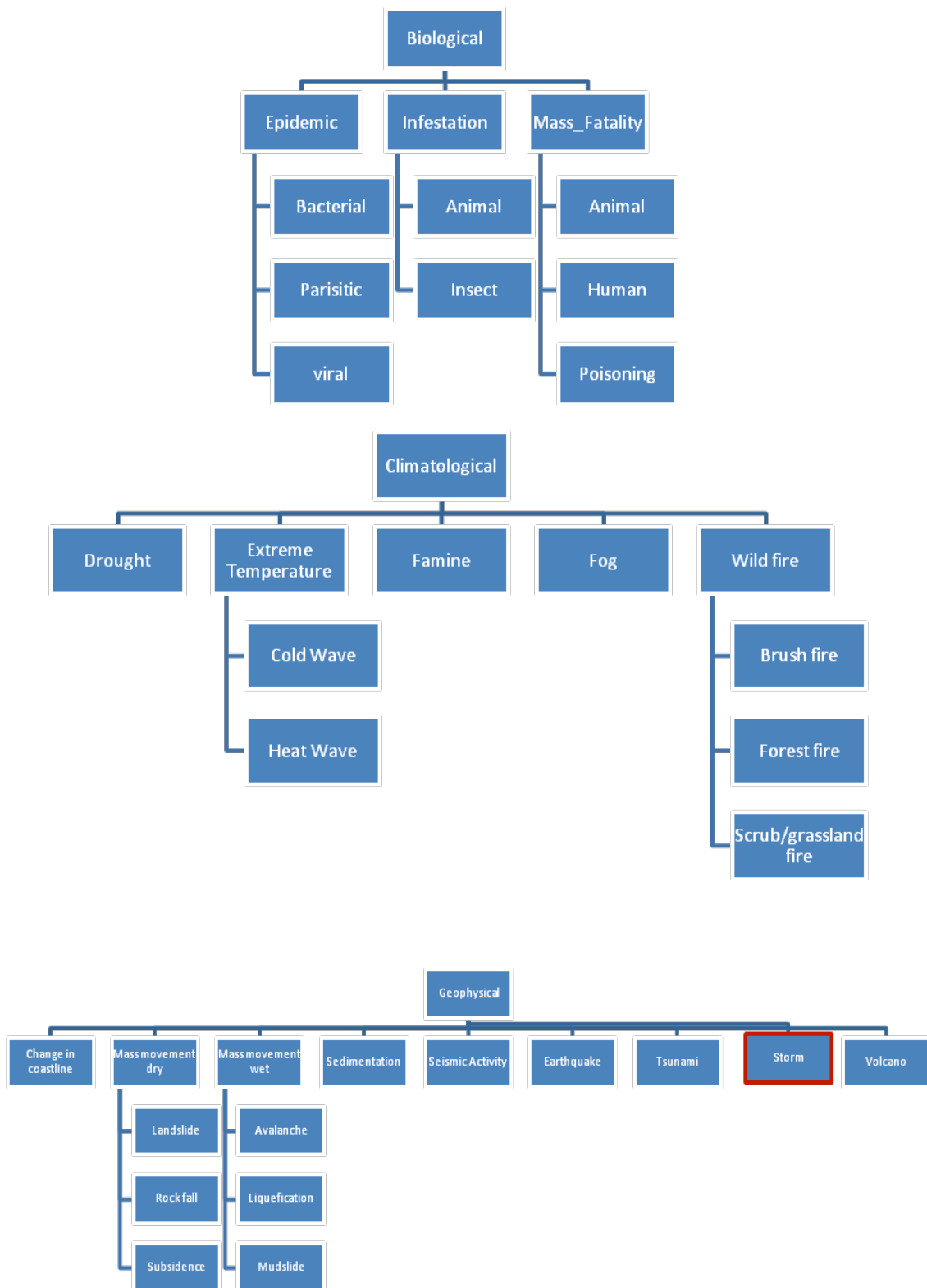
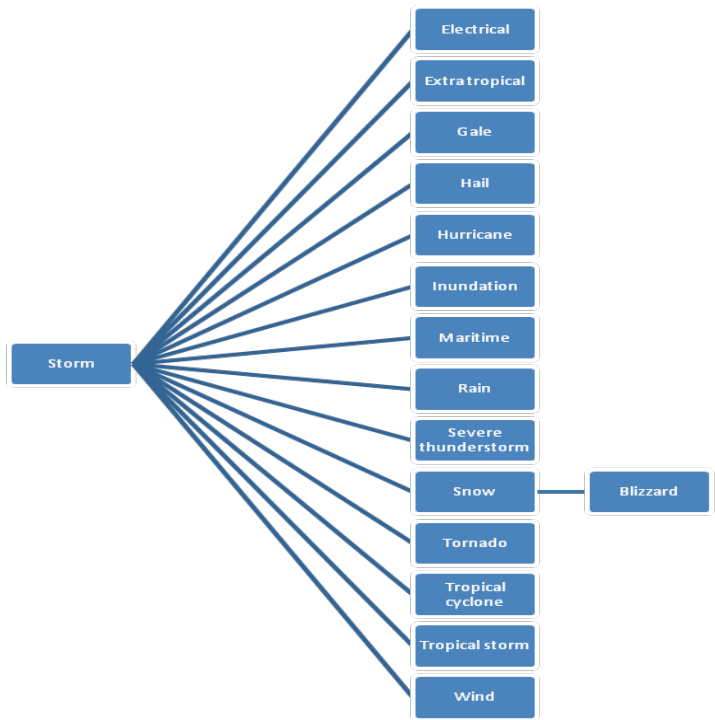**Figure 2: Sublevels of Natural disasters.**

Figure 3: Bias toward storm induced disasters (more sublevels for greater granularity).
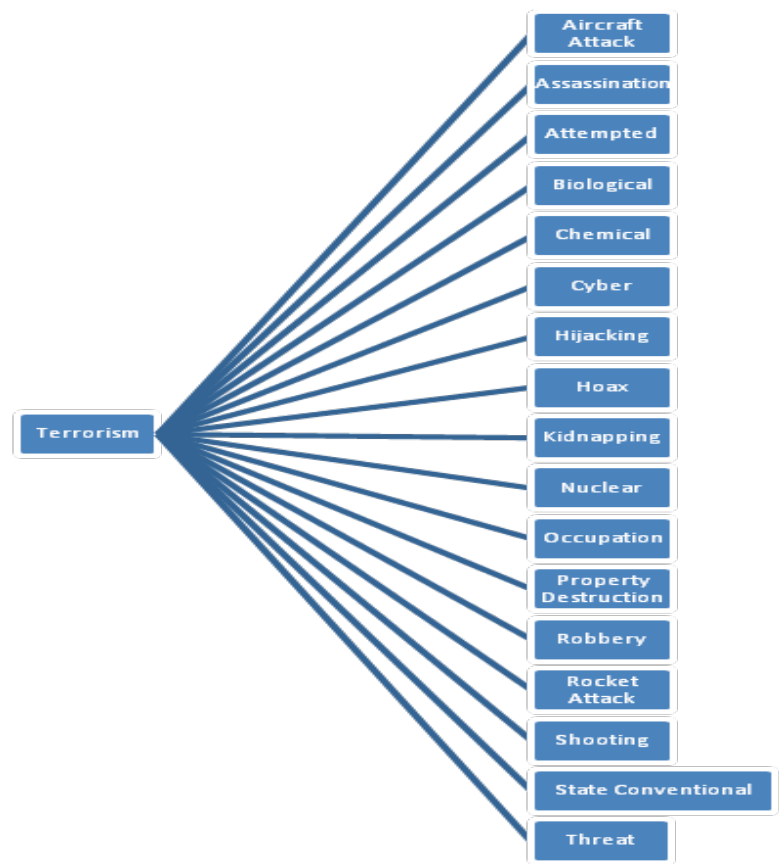

Figure 4: Bias toward Terrorism events.

We are currently mapping the attributes or slots captured by the databases, e.g., the number of people impacted or the number of homes destroyed, for a disaster into this draft ontology. Initial efforts show that a comprehensive representation of and record of facts related to disasters will be captured by associating a disaster with the appropriate elements and obtaining values for their attributes. One attribute related issue remaining to be resolved is geographic location. While it may seem that all disaster database systems would simply adopt a standard GPS representation that could be translated to other coordinate systems as desired, this is not the case. Although several databases included such attributes, e.g., longitude and latitude, they all include other, mostly unique, geographic representations including states, provinces, counties, boroughs, etc., apparently intended to allow for the maximum flexibility of definition. This is also complicated by the nature of some disasters that have impacts across multiple countries, states within countries, and cities.

CONCLUSION

Our preliminary findings reveal the nature of the focus of existing databases of disasters provide as a place for beginning to develop an ontology for our Digital Library intended to allow people involved in the disaster to preserve their experiences and provide researchers with the ability to use the information effectively. As the classification types of additional databases are merged into the ontology we expect that fewer and fewer new elements will be added.

Once the final ontology is developed and its usefulness confirmed, we will apply it in various ways. For example, we will classify resources to suit the differing needs of a variety of stakeholder groups, using the CTR ontology [12]. We will use concepts and their relations in the ontology to monitor social media. For example, we will be notified when a disaster event happens and people begin to communicate about it on the social network; that will guide our re-tweeting to suitable community and neighborhood groups.   Semantic browsing and query expansion will be provided, too. They should lead to increased user satisfaction with the information retrieval capabilities of CTRnet. To further ensure scalability, and to move toward sustainability, the CTR ontology will be moved to a public space to ensure that interested members of the community will continuously update it for broader use.

ACKNOWLEDGMENTS

REFERENCES