2011

From Glossaries to Ontologies: Disaster Management Domain

Katarina Grolinger
Western University, kgroling@uwo.ca

Kevin P. Brown

Miriam A.M. Capretz
Western University

From Glossaries to Ontologies: Disaster Management Domain

Katarina Grolinger, Kevin P. Brown, Miriam A.M. Capretz
Department of Electrical and Computer Engineering, Faculty of Engineering
The University of Western Ontario
London, ON, Canada N6A 5B9
{kgroling, kbrow43, mcapretz}@uwo.ca

***Abstract* - Our society's reliance on a variety of critical infrastructures (CI) presents significant challenges for disaster preparedness, response and recovery. Experts from different domains including police, paramedics, firefighters and various other CI teams are involved in the fast paced response to a disaster, increasing the risk of miscommunication. To ensure clear communication, as well as to facilitate CI software interoperability, a common disaster ontology is needed.**

We propose using the knowledge stored in domain glossaries, vocabularies and dictionaries for the creation of a lightweight disaster management domain ontology. Glossaries, vocabularies and dictionaries are semi structured representations of domain knowledge, where significant human effort has been invested in choosing relevant terms, determining their definitions, acronyms, synonyms and sometimes even relations. We use that knowledge built into semi formatted documents for ontology learning. In particular, we look at five glossaries/vocabularies from the disaster management domain and analyze their content similarity and structure. A lightweight disaster ontology is created exploiting the structure of the semi-structured source documents.

Keywords- Ontology, Glossary, Disaster Management, Ontology Learning

I. 1. INTRODUCTION

Today we rely heavily on a variety of critical infrastructures (CI) such as electrical systems, water supplies, telecommunications, transportation, emergency services and others. Each of these systems is highly complex with their daily operation, maintenance and repairs requiring specialized domain knowledge. On the other hand, there exists significant interdependency between these systems: water distribution systems rely on electricity to power water pumps and emergency response teams rely on a variety of telecommunication methods. Some of the interdependencies are relatively simple, such as the two previously mentioned, while the others are emergent behavior of system of systems. The significance of the cascading effects caused by interdependencies of CI has been recognized and significant efforts have been made in attempts to understand and manage them better [1] [2].

The variety of infrastructure systems involved, together with their interdependencies, demand an interdisciplinary approach to disaster management. Involvement of experts from different domains, and the need for exchange of information across domains (between people, as well as between machines), represent significant challenges in achieving successful communication. A word in common to two or more domains may have different meanings; or, conversely, different terms may represent the same concept. In fast paced emergency response situations, this may cause misunderstandings and possibly result in severe consequences.

The importance of having a common understanding within the disaster management field has been recognized; consequently, different glossaries, vocabularies and dictionaries have been created by multiple agencies involved in the disaster prevention, response and recovery processes [3] [4] [5] [6] [7]. Depending on the main focus of the agency, these glossaries, vocabularies and dictionaries vary with regards to the terms they include as well as term definitions. Furthermore, because they are mainly created to be used by people, they are in text form (PDF or text files) that cannot be easily read and understood by computers. As such, these glossaries, vocabularies and dictionaries cannot be used in the information systems that are becoming an integral part of any disaster response process. The way of making this knowledge available for use by computers is through ontologies.

An ontology is a formal, explicit specification of a shared conceptualization [8] that provides a common understanding of information. Additionally, ontologies provide a way of representing human knowledge, making it readable and understandable for machines. This, in turn, represents the basis for achieving semantic interoperability.

In this paper we look at five glossaries and vocabularies each from a different Canadian or American disaster management agency, analyze their differences and propose the creation of an ontology from these glossaries, vocabularies and dictionaries. In the remainder of the paper, the term glossary will be used to represent glossaries, vocabularies and dictionaries.

This paper is organized as follows: Section II reviews related work including ontologies in disaster management and ontology learning. Emergency management glossaries are presented in Section III together with comparative analyses.

Section IV describes ontology creation from glossaries and the conclusions and future work are presented in Section V.

II. Related Work

The main directions of research efforts related to CI interdependencies are: research dealing with analyzing past incidents that involved CI interdependencies [1] [2], studying infrastructure interdependencies through the use of simulators [9] [10] [11] and ontologies in disaster management [12] [13]. Particularly relevant to our work is the work on ontologies.

Peng et al. [12] propose the Emergency Case Ontology Model (ECOM) as a way of organizing the emergency case knowledge by taking into consideration relations among emergency cases. The proposed model handles heterogeneity among different earthquake disasters, but it is earthquake specific and does not consider other emergencies.

Castorini et al. [13] propose the Knowledge Base System (KBS) founded on ontologies with the main goal of modeling CI and their interdependencies. The proposed framework consists of: MKIONT (Meta Knowledge Infrastructure ONTology) which defines a template for conceptualization; IONT (Infrastructure ONTology) which represents knowledge of a specific CI domain (e.g. water distribution or telecommunication); FONT (Federation ONTology) which describes interactions between infrastructures and the Gateway which provides a connection between the KBS and domain simulators.

Another field relevant to this work is ontology learning, which is the process of building an ontology from scratch by enriching, or adapting an existing ontology in a semi-automatic fashion using distributed and heterogeneous knowledge and information sources. The majority of research efforts in ontology learning focuses on learning from unstructured sources [14] [15] [16], primarily text documents, as a significant portion of today's knowledge is stored in such form. Gómez-Pérez and Manzano-Macho compared different ontology leaning methods from unstructured text [17], and presented advantages and disadvantages of each method. Other possible sources of information for ontology learning are structured sources such as databases and semi-structured documents including XML schemas, web pages, glossaries, dictionaries, glossaries and vocabularies. Since semi-structured sources have various structure elements, the structure can be exploited to extract concepts and/or relations. Domain glossaries, dictionaries and vocabularies extract significant terms from the domain, and often contain synonyms, acronyms and related terms as well.

Zhao and Li [18] propose ontology learning from the hierarchy structure of organization websites. The approach is motivated by the observation that the organization web site is organized in a hierarchical sitemap that reflects a shared view of the organization structure. Consequently, they use the sitemap hierarchy to create the lightweight organizational ontology.

Karoui et al. [19] combine exploiting the structure of the HTML documents with natural language processing techniques for ontology learning. They propose the Contextual Concept Discovery (CCD) algorithm based on K-means clustering and guided by a structural context. In HTML documents they observe physical links, such as heading-paragraph links that represent the structure of documents and logical links that represent links between tags, such as keyword tags. When terms appear in the same context, within the same block tag or within linked elements, it indicates their co-relation. This structural context drives the incremental use of K-means algorithm in identification of ontology concepts.

Davulcu et al. [20] use taxonomy-directed web sites to bootstrap the ontology population task of extracting instances of concepts and their classification into ontology concepts. OntoMiner detects HTML markup and turns it into hierarchical structures that are in turn used for ontology population. Shinzato and Torisawa [21] use itemization and listing in HTML documents to extract hyponym relations.

Navigli and Velardi [22] enrich the CIDOC CRM cultural heritage core ontology using the Art and Architecture Thesaurus (AAT). Descriptions of the meanings of the terms from AAT are processed using NLP (Natural Language Processing) techniques, annotated with CIDOC properties and formalized. The core ontology is enriched from formalized term definitions.

III. Emergency Management Glossaries and Vocabularies

Glossaries and vocabularies play a significant role in emergency management due to the importance of clear communication during disaster response. Misunderstandings could lead to severe consequences, even the loss of life. Therefore government and private agencies involved with disaster management often create and publish dictionaries of relevant emergency terms. Depending on the focus of each agency, the included terms and their definitions may be significantly different.

We have analyzed five glossaries/vocabularies readily available from the web. Two are from Canadian sources: the Emergency and Crisis Communication Vocabulary from Government Services Canada [3] and the EMO (Emergency Management Ontario) glossary [4] from the Ontario provincial government. The remaining three sources are American: NIMS (National Incident Management System) glossary [5], ICDRM (The Institute for Crisis, Disaster and Risk Management) glossary [6] and ICS (Incident Command System) glossary [7]. The five listed glossaries were chosen since they are relatively generic nonspecific management glossaries dealing with generic disasters and are not disaster (i.e. flood or earthquake) or responder type (i.e. firefighters or CI teams) specific. The ICS glossary is somewhat specific since it deals with command and control in particular, but its main goal is coordination among the different actors in emergency situations which is not highly dependent on disaster or responder type. The two Canadian sources contain English and French terms and definitions, but for the purpose of this analysis, due to the need to compare to American glossaries which are only available in English, we have considered only the English terms of the glossaries. The same approach can be applied on the French part of those glossaries, or even glossaries in different languages.

A. Glossary Content Comparison

All five observed glossaries describe the same domain and therefore it is expected that the terms included in them are similar and that a high number of terms is defined in most glossaries. Some discrepancy between Canadian and American dictionaries may exist due to the slight differences of American and Canadian English. As a first step, we looked into the five glossaries to see if, and to which degree, our expectations of content similarity were correct.

From the five observed glossaries, four are relatively close in the size (Table I) and include between 115 and 167 term definitions, while ICDRM is significantly larger, including 572 term definitions. As opposed to the other four glossaries that are intended to be used in practice, ICDRM is established for the purposes of emergency management education and practice. ICDRM was created by the Institute for Crisis, Disaster, and Risk Management, at The George Washington University.

Table II depicts the overlap between the observed glossaries. Even though all five glossaries are relatively generic emergency glossaries, there is very little overlap between them. Only seven terms are defined in all five glossaries: *emergency, hazard, mitigation, preparedness, recovery, response* and *threat*. And only five terms appear in four glossaries: *communications, incident, incident management team, prevention* and *public information officer*. Some terms that could be considered significant disaster management terms, such as *risk assessment, disaster, crisis* and *alert*, are defined in only two glossaries. This demonstrates that a single glossary does not fully cover the domain and cannot be used as a standalone source for the creation of a disaster ontology. The high number of terms appearing in only one glossary is in part caused by the fact that the ICDRM glossary contains significantly more terms than the remaining four.

To further investigate the commonalities among glossaries, we observed the number of overlapping terms between pairs of glossaries. Table III shows the number of overlapping terms in pairs of glossaries and the percentage of relative overlap. The relative overlap is calculated as:

```
Relative overlap = (2 x number of overlapping
terms) / total number of terms in the pair of
glossaries x 100%.
```

The overlap between the Canadian (Emergency and Crisis Communication and EMO) and the American glossaries is relatively low – 10% or less. However, overlap between the two Canadian glossaries is still only 14%. A high number of term overlap occurs between the different American glossaries, with the highest between NIMS and ICS at 52%. This high term overlap can be explained by the fact that both ICS and NIMS are created by Federal Emergency Management Agency (FEMA).

TABLE I. GLOSSARY SIZE

	Emer. and Crisis Com	EMO	NIMS	ICDRM	ICS
# of terms	115	129	167	572	153

TABLE II. TERM OVERLAP BETWEEN GLOSSARIES

Number of glossaries term appears in	Number of terms
1	638
2	84
3	67
4	5
5	7

TABLE III. NUMBER OF OVERLAPPING TERM IN PAIRS OF GLOSSARIES AND RELATIVE OVERLAP.

	Emer. and Crisis Com	EMO	NIMS	ICDRM	ICS
# of terms	115	129	167	572	153
Emer. and Crisis Com		17 (14%)	13 (9%)	30 (9%)	10 (7%)
EMO			14 (9%)	34 (10%)	9 (6%)
NIMS				93 (25%)	84 (52%)
ICDRM					81 (22%)

There is a significant overlap between NIMS and ICDRM where 25% of the overlap is caused by the fact that ICDRM is developed with NIMS as its basis; moreover, a high number of ICDRM terms cite NIMS' definitions.

Among the five glossaries, we expected significant term overlap due to their shared domain and purpose, but the findings are on the contrary. Only high overlap is found between glossaries created within the same agency, ICS and NIMS glossaries. Overlaps between other pairs are generally low, and somewhat higher between pairs of glossaries from American sources.

Even for terms defined in all glossaries, definitions are often quite different across the glossaries; this is illustrated in Table IV with *Threat* as an example. In the five glossaries, there are five different definitions of term *Threat*. ICDRM gives two definitions of the term *Threat*, with one being the same as the definition in ICS.

In some situations, different terms have similar meanings. If we look at definition of the term *Hazard* in the ICDRM and NIMS glossary, 'Something that is potentially dangerous or harmful, often the root cause of an unwanted outcome,' it is very similar to the definition of the term *Threat* in the EMO glossary. The EMO glossary defines a *Threat* as, 'A person, thing or event regarded as a likely cause of harm or damage.' The two terms, *threat* and *hazard*, are defined as distinct terms in all five glossaries, but some of their definitions make it hard to distinguish between the meanings.

TABLE IV. TERM *THREAT* IN DIFFERENT GLOSSARIES

Glossary	Term Definition
Emerg. and Crisis Comm.	The combination of the presence of a hazard and an exposure pathway.
EMO	A person, thing or event regarded as a likely cause of harm or damage.
NIMS	Natural or manmade occurrence, individual, entity, or action that has or indicates the potential to harm life, information, operations, the environment, and/or property.
ICDRM	An indication of possible violence, harm, or danger.
ICDRM	The possibility of a hazard occurrence; something that has the potential to cause harm.
ICS	An indication of possible violence, harm, or danger.

This analysis shows that even though the five glossaries deal with the same domain and have the same purpose, they are very different in the terms that they define as well as in term definitions. Therefore, the creation of a domain ontology needs to use a variety of domain glossaries to encompass vocabularies of different domain members and to achieve better domain coverage.

B. Glossary Structures used for Ontology Creation

Glossaries, as semi-structured documents, have structure and formatting that can be used to facilitate ontology creation. The typical structure of a glossary is a term or label followed by the term's definition. For example, '**Communications:** The process of transmission of information through verbal, written, or symbolic means.' In all five observed glossaries, the term label is distinguished from the rest of the text by bold font; for example, '**Hazard:** Something that is…'. In three sources, the term is separated from its definition by a colon, while the other two use new lines. Glossary terms are concepts significant for a domain of interest. Therefore, we extract glossary terms using source document formatting and create initial ontology concepts.

Acronyms are typically included in glossaries. In [3], acronyms are separated from the term by using a semicolon: '**business resumption planning; BRP**'. In the remainder of the observed glossaries, acronyms are in brackets following the term: '**Emergency Operations Center (EOC):**…'. This is used to extract properties of the concepts for the ontology. Some of the documents, such as EMO, NIMS and ICS also have a separate section for lists of the acronyms, where only the acronym and its meaning are listed. Often, this is a duplication of the acronym listed with the term definition.

Redirection is commonly used to lead from one term to another one: '**Action Plan:** See Incident Action Plan'. If the term does not have definition, but it only has redirection, redirection is used to lead to a synonym term where the description is specified. In the observed documents, redirection is performed through the use of the 'See' word preceding a redirecting term. Therefore, if the term does not have a description and it is followed by a redirecting 'See' word, the two terms are considered synonyms. For the identification of synonyms, EMO also uses the 'synonym' word in formatting such as '**full-scale exercise** (synonym: **field exercise**)'.

Some terms are described and also contain a redirecting 'See' word; for example: '***Competency***: A specific knowledge element… See "Proficiency"'. In this case, description of a redirecting term c*ompetency* and a redirected *proficiency* term are not the same, and therefore the terms are not synonyms. In this situation 'See' indicates a related or similar concept. In the presented example, 'See' indicates that term c*ompetency* is related to term *proficiency.* EMO uses 'See also' in lieu of 'See' from the presented example to indicate related terms: '**Incident Action Plan (IAP):** An oral or written plan ... See also "Action Plan."'. These two patterns are used for the creation of relations between ontology concepts. In the Emergency and Crisis Communication Vocabulary, relations among terms are given more significance than in the other observed documents. The abbreviation 'cf.' is used to identify a cross-reference to a related concept, each being separated by a semi-colon; for example, '**mitigation** … cf. emergency management; preparedness; recovery; response; resumption'. Of the 115 terms defined in this document, related term(s) are specified for 91. Commonly, several related terms are specified for a single term, bringing the total number of relations specified in this way to 242. Even though this pattern appears in only one of the observed glossaries, it is a significant resource for the creation of relations in an ontology.

Some glossaries distinguish among different meanings of a single term. The emergency and Crisis Communication Vocabulary uses number superscripts to indicate different meanings, while EMO uses numbers in brackets following the term: '**hazard (1)** A risk that is a threat. **hazard (2)** An event ...'. ICDRM uses bullets to specify different definitions, as in:

'**Hazard**:
- A potential or actual force, …
- Something that is …'

ICDRM has formatting features indicating taxonomic hierarchies. For example, the term *volunteer* is in the ICDRM, but its definition is not specified. The term *volunteer* is followed by bulleted list where each bullet specifies definitions of a specific kind of volunteer, such as *Accepted volunteer*, *Affiliated volunteer*, *Recruited volunteer* and others. This pattern defines a hyponym (is-a) relation where one concept is a subconcept of another concept; for example, *Affiliated volunteer* is a special kind of *Volunteer*. This is the only structural pattern in observed glossaries that we use for the extraction of hyponym relations.

IV. DISASTER ONTOLOGY

The structures of the glossaries represent a strong foundation for the creation of an initial domain ontology. Domain experts' knowledge was used in the process of the glossary creation and it is built into the glossary itself. Terms defined in the glossary are identified as relevant terms by the people and organizations that created the glossary. Other elements, such as synonyms, related terms, acronyms and subconcepts, are extractions of domain knowledge as well.

We use the formatting aspects of glossaries described in subsection III.B as a source of information for the ontology's creation. We use only the formatting elements to extract an ontology, without the use of any natural language processing techniques. Fig. 1 depicts the use of the formatting elements of glossaries in ontology learning. The left column shows the structural elements observed in one or more glossaries. It is followed by the example of each pattern from one of the observed glossaries. Those formatting patterns identify fragments of the glossaries that are translated into different ontology elements as shown in the ontology element column of the Fig. 1. Patterns used in the ontology creation process are only those found in the five disaster management dictionaries analyzed. The use of other glossaries, dictionaries or vocabularies may demand different or additional patterns depending on the structure of the source document.

Using a single glossary for ontology creation would limit the ontology to the view of the domain described by the glossary creator; this single view is likely to not be shared by other participants in the same domain. The coverage of the domain would also be limited. To alleviate this, we use the multiple glossaries described in Section III for the ontology creation.

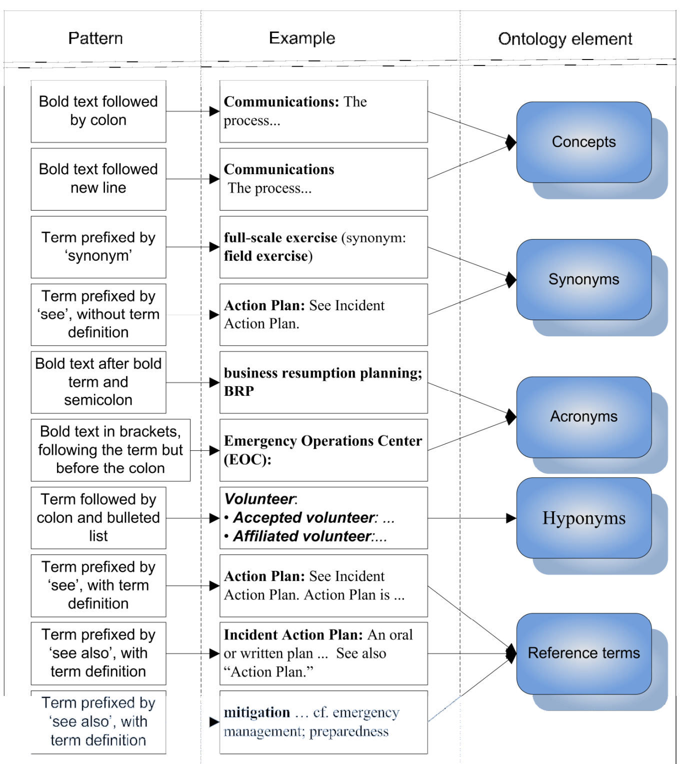

Figure 1. From formatting element to ontology component

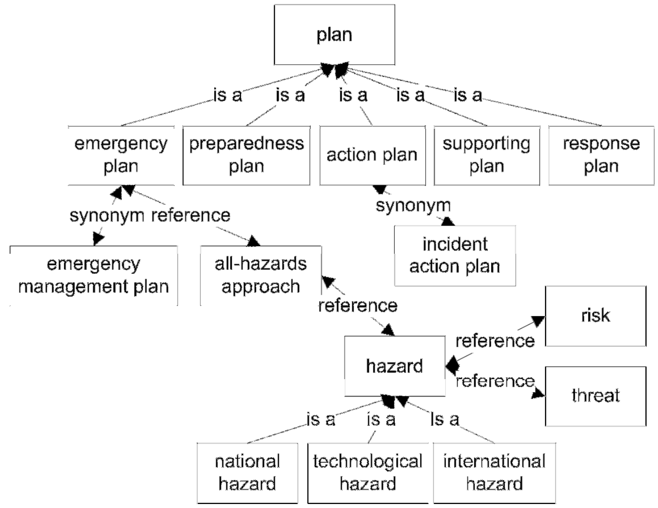

Figure 2. Fragment of ontology created from glossaries

Fig. 2 illustrates a fragment of the ontology created from the five observed glossaries using only the formatting elements of the source documents without the use of any natural language processing techniques. Concepts and relations are extracted from different glossaries. The concepts *hazard* and *threat* appear in all five glossaries. The concept *Plan* is defined only in the ICDRM glossary, while the other glossaries contain definitions of more specific plans, such as *preparedness plan* and *response plan*. *Emergency management plan* is identified as a synonym of *emergency plan* and *action plan* is a synonym of *incident action plan*.

For use by machines and software systems, the created ontology can be represented in an ontology language of choice, such as OWL (Web Ontology Language), OIL (Ontology Interchange Language) or others. The choice of representation language does not change the ontology learning process, but it only changes how the ontology is represented for automatic processing.

V. CONCLUSIONS AND FUTURE WORK

Emergency preparedness, mitigation, response and recovery involve response teams, various CI teams, police and others. Involvement of different views of the same domain presents a significant challenge in achieving a successful communication process. Ontologies are a tool for achieving a common understanding of the domain as well as for facilitation of software interoperability.

We explore the use of glossaries as semi-formatted stores of domain knowledge for ontology creation. Domain experts' knowledge was used in the process of glossary creation and is built into the glossary in the form of content and formatting. Terms chosen to be included and defined in a glossary are identified as relevant terms for the domain by domain experts and organizations involved in the glossary creation. We exploit this knowledge built into glossaries for ontology learning.

Specifically, we observe five glossaries from the disaster management domain and analyze their formatting, content and overlap. Even though they all represent relatively generic emergency management glossaries, their content overlap is very low, with only five terms appearing in all five glossaries. Content is only similar, with term overlap of 52%, between two glossaries created by the same emergency management agency. This indicates that for ontology creation, it is preferable to use multiple domain glossaries from different sources. This will lead to better domain coverage and facilitate a true shared conceptualization.

Formatting of the documents is similar across observed glossaries. All five glossaries use similar formatting to distinguish between terms and their definitions, and similar methods to identify synonyms, acronyms and related terms. We use the formatting of documents to extract terms, synonyms, acronyms, hyponyms and referenced terms to create an initial ontology. The main advantage of this approach is the use of domain knowledge built into domain glossaries and the relative simplicity of the processing. The initial ontology created using this approach is lightweight, without considerable detail, but the quality of included terms and relations is high due to the high reliability of the source document. The created lightweight ontology can be enriched by applying further processing using statistical methods or natural language processing methods, such as the approach proposed by Navigli and Velardi [22].

The direction of the future work is towards fully utilizing the structure of semi-formatted documents for ontology learning. A rule engine that will enable specifying custom rules for the extraction of concepts and relations from generic semi-formatted documents needs to be created. This rule engine will enable the user to specify how formatting should be used in ontology learning. Because this approach creates lightweight ontologies, we want to integrate it with other ontology leaning mechanisms from un-structured text. Also, a way of distinguishing between the significance of different source documents is needed; that is, a method that will account for source relevance and reliability.

Acknowledgment

Support for this work was provided by Canada's Advanced Research and Innovation Network (CANARIE) and Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] S.M. Rinaldi, J.P. Peerenboom, T.K. Kelly. “Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencie,” *Control Systems Magazine, IEEE,* vol. 21, Issue 6, 2001, pp. 11-25.

[2] E. Luiijf, E. Nieuwenhuijs, M. Klaver, M. van Eeten, E. Cruz, Empirical Findings on Critical Infrastructure Dependencies in Europe, Critical Information Infrastructure Security, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2009, pp. 302-310.

[3] Emergency and Crisis Communication Vocabulary, Public Works and Government Services Canada, http://www.btb.gc.ca/publications/documents/crise-crisis.pdf.

[4] Emergency Management Ontario (EMO) English-French Glossary, Emergency Management Ontario, Ministry of Community Safety and Correctional Services, http://www.onterm.gov.on.ca/EMOlexicon.pdf.

[5] National Incident management System (NIMA) glossary, Federal Emergency Management Agency, United States Department of Homeland Security, http://www.fema.gov/emergency/nims/Glossary.shtm.

[6] ICDRM/GWU Emergency Management Glossary of Terms. The Institute for Crisis, Disaster, and Risk Management (ICDRM) at the George Washington University (GWU), Washington, D.C., http://www.gwu.edu.proxy2.lib.uwo.ca:2048/~icdrm/publications/PDF/EM_Glossary_ICDRM.pdf.

[7] Incident Command System (ICS) glossary, Federal Emergency Management Agency, United States Department of Homeland Security, http://training.fema.gov/EMIWeb/IS/ICSResource/assets/ICSGlossary.pdf.

[8] R. Studer , V.R. Benjamins, D. Fensel. “Knowledge Engineering: Principles and Methods,” *Data and Knowledge Engineering,* vol. 25, Issue 1-2, 1998, pp. 161-197.

[9] H.A. Rahman, M. Armstrong, D. Mao, J.R. Marti, “I2Sim: A Matrix-Partition Based Framework for Critical Infrastructure Interdependencies Simulation,” *Proceedings of the Electric Power Conference,* 2008, pp. 1-8.

[10] A. Usov, C. Beyel, E. Rome, U. Beyer, E. Castorini, P. Palazzari, et al., “The DIESIS Approach to Semantically Interoperable Federated Critical Infrastructure Simulation,” *Proceedings of the Second International Conference on Advances in System Simulation,* 2010, pp. 121-128.

[11] A. Tofani, E. Castorinia, P. Palazzaria, A. Usovb, C. Beyelb, E. Romeb, et al. “Using Ontologies for the Federated Simulation of Critical Infrastructures,” *Proceedings of the International Conference on Computational Science,* vol. 1, Issue 1, 2010, pp. 2301-2309.

[12] Y. Peng, W. Wenjun, D. Cunxiang, “Application of Emergency Case Ontology Model in Earthquake,” *Proceedings of the International Conference on Management and Service Science,* 2009, pp. 1-5.

[13] E. Castorini, P. Palazzari, A. Tofani, P. Servillo, “Ontological Framework to Model Critical Infrastructures and their Interdependencies,” *Complexity in Engineering,* 2010, pp. 91-93.

[14] J. Völker, P. Haase, P. Hitzler, “Learning Expressive Ontologies,” *Proceedings of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge,* vol. 167, Issue 2/8/2010, 2008, pp. 45-69.

[15] A. Maedche, S. Staab, “Discovering Conceptual Relations from Text,” *Proceedings of the 13th European Conference on Artificial Intelligence,* 2000, pp. 321-325.

[16] M.A. Hearst, “Automatic Acquisition of Hyponyms from Large Text Corpora,” *Proceedings of the 14th conference on Computational linguistics,* vol. 2, 1992, pp. 539-545.

[17] A. Gómez-Pérez, D. Manzano-Macho. “An Overview of Methods and Tools for Ontology Learning from Texts,” *The Knowledge Engineering Review,* vol. 19, Issue 3, 2004, pp. 187–212.

[18] Y. Zhao, J. Li, “Domain Ontology Learning from Websites,” *Ninth Annual International Symposium on Applications and the Internet,* 2009, pp. 129-132.

[19] L. Karoui, M. Aufaure, N. Bennacer, “Contextual Concept Discovery Algorithm,” *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference,* 2007, pp. 460-465.

[20] H. Davulcu, S. Vadrevu, S. Nagarajan, I.V. Ramakrishnan, “OntoMiner: Bootstrapping and Populating Ontologies from Domain-Specific Web Sites,” *Intelligent Systems, IEEE,* vol. 18, Issue 5, 2003, pp. 24-33.

[21] K. Shinzato, K. Torisawa, “Acquiring Hyponymy Relations from Web Documents,” *Proceedings of North American Chapter of the Association for Computational Linguistics - Human Language Technologies NAACL-HLT,* 2004, pp. 73-80.

[22] R. Navigli, P. Velardi, “From Glossaries to Ontologies, Extracting Semantic Structure from Textual Definitions,” *Proceeding of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge,* vol. 167, 2008, pp. 71-104.