Automated unsupervised ontology population system applied to crisis management domain

Yohann Chasseray, Anne-Marie Barthe-Delanoë, Stéphane Négny, Jean-Marc Le Lann

Automated unsupervised ontology population system applied to crisis management domain

Yohann Chasseray*
Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France†
yohann.chasseray@ensiacet.fr

Anne-Marie Barthe-Delanoë
Centre Génie Industriel, Université de Toulouse, IMT Mines Albi, France‡
anne-marie.barthe@mines-albi.fr

Stéphane Négny
Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France
stephane.negny@ensiacet.fr

Jean-Marc Le Lann
Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France
jeanmarc.lelann@ensiacet.fr

ABSTRACT

As crisis are complex systems, providing an accurate response to an ongoing crisis is not possible without ensuring situational awareness. The ongoing works around knowledge management and ontologies provide relevant and machine readable structures towards situational awareness and context understanding. Many metamodels and ontologies have been designed to collect and organize crucial information for Decision Support Systems. However, feeding metamodels is still an issue (limitation of manual instantiation, domain centered methods). The next challenge into crisis management is to provide tools that can process an automated population of these metamodels/ontologies. The aim of this paper is to present a strategy to extract concept-instance relations in order to feed crisis management ontologies. The presented system is based on a previously proposed generic metamodel for information extraction and is applied in this paper to three different case studies representing three different crisis namely Ebola sanitarian crisis, Fukushima nuclear crisis and Hurricane Katrina natural disaster.

Keywords

Automated knowledge extraction, crisis management systems, ontologies, experience feedback exploitation, background knowledge acquisition.

INTRODUCTION

Situational awareness (Endsley 2001) is crucial to manage complex systems, especially crisis situations. Understanding context (with respect to time and space) is critical for decision-making. Indeed, Decision Support Systems rely on the availability and quality of the information describing the context. This information is usually provided by sensors, live reports, etc. (feeding the Common Operational Picture during the response stage) as well as feedback from past crisis (especially at the preparation stage).

With the rise of the interconnection and communication between objects and people, the so-called Internet of Things (IoT) (Kopetz 2011) and initiatives like crowd sourcing and citizen participation (Ofli et al. 2016; Castillo 2016), multidisciplinary open archives (HAL, arXiv to mention a few) and Open Data, we are facing an exploding amount

*corresponding author
†https://lgc.cnrs.fr/
‡http://www.mines-albi.fr

of emitted data. This tremendous stream of raw data about the environment, people, organizations, processes, activities, policies, past events, etc., known as Big Data (Demchenko et al. 2013) can be used as an input for ensuring situational awareness. It can be valuable in the context of crisis management as it often contains a huge volume of unstructured knowledge. However, as this knowledge is unstructured, it is not actionable as Decision Support Systems cannot automatically process it.

To bridge this gap between data and actionable knowledge, meta-modeling and ontology have gained popularity in data-intensive domains in the last decade (Skvortsov et al. 2016; Kontopoulos et al. 2017). They provide languages (and first-order logic for the later) to organise and structure domain knowledge in a way that it can be interpreted by Decision Support Systems. In crisis management area, numerous meta-models and ontologies exist, structuring either specific kind of crisis (e.g. floods (Fertier et al. 2020), fire emergency situations (Bitencourt et al. 2018)) or the global system at one or several stages of the crisis lifecyle (e.g. crisis situation representation (Benaben et al. 2020), response planning (Gaur et al. 2019)).

Knowledge bases (i.e. populated ontologies) constitute a condensed, structured and machine readable abstraction of the knowledge of a domain. Thus using this knowledge as a resource in Decision Support Systems is a way to provide knowledge that was not necessarily available at first sight for decision makers. In these terms, using ontologies, populated into knowledge bases from external data sources, participates in situational awareness as it allows a Decision Support System to have a better understanding of the ongoing situation.

However, instantiating meta-models and populating ontologies with massive volumes of data remains a challenge. Manual instantiation and population are no longer an option, as it is not only error-prone and time-consuming (especially in a crisis management context) but unrealistic regarding the variety of data sources and the amount of produced data. In addition, existing population systems usually either required structured data or manually annotated documents, or were designed for a given knowledge structure. These statements go in favor of the automation of ontology population in a generic way. Automating ontology population would indeed allow Decision Support System to have access to many more sources of information that could not have been processed manually in a reasonable amount of time.

Two main issues can then be identified in the field of ontology population being (i) manual population that is unsuitable considering the huge amount of available data and (ii) the limitation to specific use cases when it comes to automatically populate an ontology. Both issue may especially impact crisis response as ontology population has to be quick in such a context and certainly not limited to few types of crisis or domain.

The purpose of this paper is then to address these issues of ontology population through the proposition of an unsupervised domain independent extraction architecture and its application to the case of crisis management. The paper is organised as follows. Section 2 explores previous related work about automated ontology population. Section 3 presents the architecture of the proposed ontology population system and the different steps of ontology population. Section 4 presents the application on a crisis management ontology and Section 5 discusses the extraction results of the case study.

RELATED WORK

As explained, the role of an ontology is predominant if one want to ensure situational awareness, for complex systems in general and in the context of a crisis in particular. The major limiting aspect to the application of ontologies is their lack of real world tangible knowledge. As many studies tackle the question of ontology population, they can be classified referring to three different aspects, which are (i) the genericity regarding the domain, (ii) the automated or semi-automated nature of the extraction process and (iii) the type of data sources covered. This section presents related work regarding these aspects.

Semi-automated to automated extraction systems

Even if the trend is to reduce the use of supervised methods (Zhang and Lu 2019), state of the art tools are still far from fully unsupervised knowledge extraction systems, especially when dealing with unstructured data. As stated by (Konys 2019), most extraction systems suppose a strong implication of domain or ontology experts. Some systems, such as Karkaletsis et al. (2006), are based on both previous data annotation and expert support to learn additional knowledge.

Some attention has also been given to the automated population of ontologies into knowledge bases through alignment with existing ontologies or even knowledge bases (Alobaidi et al. 2018; Roth-Berghofer et al. 2010). This kind of approach reuses already gathered knowledge and reinvest it to support further knowledge extraction. Once again, it assumes that some structured knowledge is already available which might be the case for medical domain but not necessarily in other domains such as crisis management.

Multi data types extraction systems

The variety of sources is a dimension that should not be avoided when harvesting big data. Most systems are developed for raw unstructured text mining. Some work are dedicated to other group of data sources such as XML semi-structured data (Tekli et al. 2016; Swaraj and Manjula 2016), web data (Jannach et al. 2009; Ferrara et al. 2014; Vicient et al. 2013), which are slightly different from raw text, but also mathematical formulas (Qin et al. 2016), databases (Liao et al. 2019; Liu and Gao 2018), or images (Huynh and Neptune 2018). However, not many studies focus on the processing of multi-source data (Remolona et al. 2017; Buitelaar et al. 2008). Remolona et al. (2017) address this issue with the HOLMES architecture welcoming data streams of different sources for the chemical domain. These approaches in return may be domain-centered and often require a huge amount of annotated data in order to be efficient.

Domain independent extraction systems

Most of the work conducted on extraction systems is guided and designed for specific case studies. The recent development of ontologies is notably led by the need of populated ontologies in the medical domain. Many work are then focused on this domain. However, there is still few extraction systems that are designed to cover multiple domains as extraction methods they use are often proper to the domain.

Approaches exist which use unsupervised methods to derive taxonomies or even ontologies from text. These approaches should mostly be associated to the field of ontology learning rather than ontology population as they create their own representation of knowledge and do not populate an existing ontology. To ensure the keeping of an existing ontology structure a generic metamodel for information extraction has been designed by Chasseray et al. (2021). This metamodel ensures the adaptability to different kind of sources and assumes the use of an existing ontology. As it describes very generic concepts it can also be adapted to many different domains. The presented system has been built on this metamodel.

METHODOLOGY AND ARCHITECTURE

The methodology initially presented by Chasseray et al. (2021) is based on a generic metamodel for information extraction. This metamodel is then integrated to a wider framework to derive a data model constituting a bridge between raw and heterogeneous data and the structured ontology that is going to be populated. The main advantage of the framework is its genericity regarding the exploited data sources and the domain described by the ontologies. In the presented framework, the ontology population is done in three main steps namely (i) instance extraction based on the ontology, (ii) expert validation and (iii) alignment of validated instances with the ontology. These three steps are detailed in the following sections.

Instance extraction : From unstructured data to information

The main task in ontology population is the research of instances (ontology individuals) that can be related to concepts (ontology classes). In the presented framework, this is done through two different processes. The first process uses a rule-based approach extracting hyponymy relations through extraction patterns coupled with natural language processing. The second process extracts terms with statistical means and uses a matching strategy based on previously extracted relations in order to derive new knowledge.

Pattern extraction

Pattern extraction has been widely used for the identification of relations in large corpora at different levels of genericity. The interest of the proposed method is its interoperability with different ontologies, whatever the domain they describe or the way they have been designed. In ontology population, the research of hyponyms is a major task as hyponymy relations represent the links between real world examples (i.e individuals of the ontology) and their conceptualised version (classes of the ontology). Thus the presented method focuses on terms extraction (potentially representing an instance) based on the hyponymy relation they might have with occurrences of ontology classes in the data.

Statistical extraction

Statistical approach has been widely used for corpora analysis from Bayesian statistics (Kunal et al. 2018) and Tf-Idf measures (Jones 1972) to Latent Semantic Analysis space and Word embedding (Altszyler et al. 2016). The aim of the statistical approach is here to select relevant candidates for further matching with already extracted instances. As candidates need basically to be terms linked to the studied domain, the choice of candidates is made comparing the frequency of appearance of terms in common language to their frequency in text.

Expert validation : from information to knowledge

The presented framework can use expert validation at the end of the extraction process. Expert validation is mainly used for evaluation of the system. The precision of the system is computed based on the expert's decisions made at this step of the process. Expert's evaluation can also help avoiding noise in extracted elements which are then used to enrich extraction tools (rules, patterns). However, it is important to specify that expert validation step occurs after the extraction of relations. This means that the human intervention and expertise is used only for evaluation purpose and only once the system already extracted knowledge. Thus, it only impacts population as it removes noise and incorrect extracted relations. The extraction process itself is considered completely unsupervised, as it does not need annotated data as input.

Still, the choice of the expert might impact the result of the evaluation. Generally, it is considered that an expert has a general knowledge of their domain providing them an accurate understanding of the concepts used in the domain ontology that is to be populated. In the specific case of crisis domain, an expert is considered at ease with the concepts of a crisis (Event, People, Measure, Actor, etc.) but being an expert in a specific type of event is not a requirement for validation.

Alignment : from the Data Model to the Knowledge Base

After relation extraction and expert validation, extracted knowledge has not yet populated the ontology as it still lies in the data model. However, as soon as the relation between an instance and its concept in the data model is confirmed as correct or declared incorrect during expert validation step, an alignment between this data model and the ontology can be processed. To drive this model transformation, alignment rules are defined at the metamodelling level between the generic metamodel and the upper ontology making each validated instance become an individual of the ontology.

CASE STUDY : ONTOLOGY POPULATION FOR THE CRISIS DOMAIN

Table 1. Concepts of the crisis represented as classes in the ontology to instantiate (based on (Benaben et al. 2020) crisis metamodel) major concepts. Some *synonyms* and core concepts (bold) are added in order to cover more vocabulary.

Partners group	Context group	Objective group	Behavior group
Actor	Good	Event	Subprocess
Service	People	Fact	**Measure**
Resource	Natural site	Crisis	**Activity**
Organisation	Danger	Intrinsic risk	**Process**
–	Emerging risk	Factor	*Indicator*
–	**Characteristic**	**Objective**	*Sensor*
–	**Threat**	*Strategy*	–
–	*Accident*	*Effect*	–
–	*Population*	*Risk*	–
–	*Area*	–	–

Crisis ontology : classes selection

As it has been presented in the introduction, we can find a lot of ontologies defined to describe the domain of crisis management. In order to build our case study and illustrate the functioning of the system, a light ontology is built based on the main concepts of crisis management. The content of this ontology, which defines generic classes for the description of a crisis are listed in Table 1. The choice is made to keep only generic concepts so that this ontology can describe any crisis. Nevertheless, the ontology population system can easily be adapted to other broader or more specific ontologies.

Generic relation extraction rules are fed by these defined classes. Thus, other terms (*italic* and **bold** terms) are added to the bunch of concepts extracted from Benaben et al. (2020) crisis response metamodel as they provide extended vocabulary for instance detection. Some of them (**bold terms**) are generic concepts extracted directly from the collaboration core metamodel (Benaben et al. 2020). Also some synonyms of the metamodel's concepts (*italic terms*) are added in order to cover more vocabulary. These added terms are still linked to crisis as they can be

Table 2. Details about analysed press and academic articles for the population of the predefined crisis ontology classes regrouped by type of crisis.

Crisis	Natural disaster (Katrina)	Nuclear crisis (Fukushima)	Health crisis (Ebola)
		Press Articles	
Number of articles		20	
First article	28th August 2005	13th March 2011	1st August 2014
Last article	26th September 2005	1st April 2011	6th May 2015
Number of words	33 269	27 903	27 985
		Academic articles and official reports	
Articles	(Murphy and Jennex 2006) (9 pages)	(Heng and Tao 2014) (5 pages)	(Landgren 2015) (7 pages)
	(Fetter et al. 2010) (5 pages)	(Guan et al. 2015) (6 pages)	(Bell 2016) (8 pages)
	(Yeletaysi et al. 2008) (8 pages)	(Segault et al. 2015) (8 pages)	(Kaner and Schaack 2016) (7 pages)
	(Boin et al. 2019) (30 pages)	(Ishigaki et al. 2015) (7 pages)	(World Health Organization et al. 2015) (27 pages)
	–	(French et al. 2017) (10 pages)	(Save the Children 2015) (40 pages)
	–	(Barthe-Delanoë et al. 2014) (5 pages)	–
	–	(OECD and Nuclear Energy Agency 2013) (60 pages)	–
Number of words	33 165	64 497	56 319

found in other knowledge management frameworks dealing with crisis management, such as the ones defined by Zschocke et al. (2010) and Calcaterra et al. (2015). Nevertheless, they are not supposed to modify the structure of the ontology containing previously defined classes.

Explored data

Within an ongoing crisis, the amount of information that can be collected must be divided into two groups of relevant information. The first group concerns the information that is specific to the context of the studied crisis. This includes real time sensor data, social networks information, institutions' announcements. The other part of information, which is also relevant for exploitation, is external to the ongoing crisis and can be gathered in documentation about previous similar crisis (scientific articles, governmental reports, etc.).

Both groups contain relevant knowledge for the population of a crisis management ontology. For example, ISCRAM Community, through the publications of past conferences, concentrates a lot of knowledge regarding crisis management. As the main conference has more than a decade of existence, there is a consequent number of papers that deal with several crisis, that occurred around the world.

The goal of this case study is then to gather a bunch of articles for automated extraction of knowledge. The chosen articles are directly related to three different types of crisis, namely natural disaster crisis (Hurricane Katrina), nuclear crisis (Fukushima nuclear disaster) and sanitarian crisis (Ebola crisis). The choice is made to select crisis of different nature in order to evaluate how the system reacts when applied to different kind of crisis. These case studies have also been chosen so that there is enough available academic and press data dealing with each crisis.

For each of these crisis, feedback reports are also added to the set of explored documents as they generally summarize particularly well the events of a crisis. For convenience the set of academic articles and feedback reports is unified under the term *academic* in the remaining of this paper. Details about the documents can be found in Table 2.

In order to illustrate its interoperability with other kinds of documents, the system is used on newspaper data as well. Textual data are then taken from news articles published online by the New York Times during periods concerning each crisis. The choice of the New York Times is motivated by the need to find a media that would provide enough data on the three different crisis. Articles are harvested thanks to queries including specific crisis relative keywords. Information about the press articles textual data can be found in Table 2. Textual data related to press articles is identified in the remaining of this paper by the term *press*.

The bias induced by the nature of media data should not be eclipsed. This is one of the reason why extraction have been tested and compared for both academic and press articles. Press articles harvesting have been limited to one media for evaluation purpose. However, such a system when used in real context should not be limited to a single source so that media bias can be at least contained, if not avoided.

Used extraction rules

In the presented case studies, only rule-based method is used to extract instances. Three extraction rules, based on (Hearst 1992) patterns, have been used to guide the extraction process. The first rule, called *X is a Y* is designed in order to detect concept-instance (C-I) relations appearing with the scheme *I is a C* or *I are C*. The second rule, called *Y such as, X1, X2* is designed in order to detect the *C such as I, I'* or *C like, I, I'* forms of concept-instance relation. The third rule is called *X = modifier Y* and tends to consider as an instance any concept that is modified by another word or group of words (called *modifier*).

Result indicators

The results of the extraction are presented aggregated by concept (i.e. classes) and with 5 indicators which are :

- The number of extracted instances for each concept (**N**) : This indicator counts all the extracted instances for each concept regardless of the decision made during validation step.
- The number of validated instances among the extracted instances (**V**) : This indicator counts all the instances that have been extracted and then validated during validation step.
- The number of validated instance with high confidence among the validated instances (**S**) : During validation step the expert can validate or invalidate instances with a different levels of confidence. Some validated instances remain very generic and can hardly be assimilated to individuals as they mostly describe concepts. These instances remain relevant as they contain some additional knowledge and may be used in different ways. However they are still not precise enough to be directly used in a Decision Support System. Hence, this indicator counts the number of instances that have been validated and considered correctly extracted and specific enough to be derived as individuals.
- Extraction precision over extracted instances (**P**) : A precision measure has been set up in order to estimate the success rate of the extraction system. As we use unlabelled text, this indicator does not provide any recall information because it is not possible to estimate the rate of missed instances. Extraction precision can be computed separately for distinct concepts, distinct types of extraction rule and also over all extracted instances. Following equation shows how precision is computed for each concept c :

$$P_c = \frac{\sum_{i \in I_c} \delta_i * \omega_i}{\sum_{i \in I_c} \omega_i} \tag{1}$$

  Where I_c corresponds to the set of extracted instances for concept c, ω_i corresponds to the confidence rate affected to the validation by the expert , and δ_i is defined as follows :

$$\delta_i = \begin{cases} 1 & \text{if instance } i \text{ has been validated} \\ 0 & \text{else} \end{cases} \tag{2}$$

- Similarity measure (**Sim**) : Since the same set of extracted instances can be validated by several different experts, the induced validations may result in different validated sets. Then it becomes necessary to monitor the rate of similarity between these validated sets. Hence, the similarity between two sets validated by expert a and expert b is measured as follows :

$$Sim_{a/b} = \frac{\sum_{i \in I} \delta_{i_{a/b}} * (1 - |\omega_{i_a} - \omega_{i_b}|)}{Card(I)} \tag{3}$$

  Where I corresponds to the set of extracted instances validated by expert a and expert b , ω_{i_a} and ω_{i_b} corresponds to the confidence rate affected to the validation respectively by expert a and expert b, and $\delta_{i_{a/b}}$ is defined as follows :

$$\delta_{i_{a/b}} = \begin{cases} 1 & \text{if experts } a \text{ and } b \text{ have both validated or both invalidated the instance } i \\ 0 & \text{else} \end{cases} \tag{4}$$

Extraction results

In this section, several use cases of the ontology population system are used to demonstrate its applicability on different types of crisis. The objective of these use cases is to detect instances that are relevant for the feeding of a Decision Support System. For each use case, the same ontology is used, containing the classes presented in Table 2, but different documents are used (see Table 2).

Tables 6, 7 and 8 present some of the extracted instances related respectively to the Fukushima crisis, Ebola crisis and Hurricane Katrina crisis. As press articles have also been analyzed, tables 3, 4 and 5 contain the results of press

Table 3. Press articles extracted relations validated by crisis expert concerning Fukushima nuclear crisis.

Concept	N	V	S	P	Examples of extracted instances		
					Instance	**Extracted data**	**Extraction rule**
Risk	17	15	**15**	0.88	increase risk [of cancer]	... *lead to an **increased risk** of cancer* ...	X = modifier Y
					radiation risk	... *the **radiation risk** to the public appears low so far* ...	X = modifier Y
					earthquake risk	... *almost all of the country lies in an **earthquake-risk** zone.* ...	X = modifier Y
Area	12	12	**11**	1.00	high - risk area	... *to strike the **higher-risk areas** southwest of fukushima* ...	X = modifier Y
					agricultural area	... *but that should not be done in **agricultural areas**, she said,* ...	X = modifier Y
					tokyo area	... *40 million consumers in the greater **tokyo area**.* ...	X = modifier Y
Accident	11	10	**8**	0.95	reactor accident	... *after the three mile island **reactor accident**.* ...	X = modifier Y
					nuclear power accident	... dangerous kind of a **nuclear power accident** because of the risk of radiation ...	X = modifier Y
					chernobyl accident	... *in the **chernobyl** nuclear **accident** of 1986* ...	X = modifier Y
People	7	6	**4**	0.83	japanese people	... *distributed thousands of pounds of food and water to the **japanese people**.* ...	X = modifier Y
					standard staffing levels	... ***standard staffing levels** [...] on the site would be 10 to 12 people* ...	X is a Y
Event	4	4	**4**	1.00	nuclear event	... *one measure of a **nuclear event** rated level 5, for example, is the melting* ...	X = modifier Y
					fukushima event	... *still, the **fukushima event** has involved a significant release of radiation* ...	X = modifier Y
Effect	3	3	**2**	1.00	health effect	... *the biggest **health effect** was cases of thyroid cancer* ...	X = modifier Y
					ill effect	... *even if people consume the water a few times, there should be no long-term **ill effects**.* ...	X = modifier Y
Danger	2	2	**2**	1.00	potential danger	... *some areas to pose a **potential danger** to people* ...	X = modifier Y
Resource	2	1	**1**	0.67	scarce resource	... *they've shared **scarce resources** of food and water.* ...	X = modifier Y
Good	1	1	**1**	1.00	dairy good	... *prohibit imports of **dairy goods** and produce from the affected region.* ...	X = modifier Y
Activity	4	1	**1**	0.25	seismic activity	... *the most violent **seismic activity*** ...	X = modifier Y
Population	1	1	**1**	1.00	fish population	... *i would definitely be monitoring **fish populations** in the area* ...	X = modifier Y
Threat	1	1	**1**	1.00	immediate threat	... *these levels do not pose an **immediate threat** to your health," mr. edano said.* ...	X = modifier Y
Service	1	1	**1**	1.00	phone service	... *a lack of **phone service** meant that they* ...	X = modifier Y
Factor	2	1	**1**	0.67	wind speed	... *based on measurements of a single factor like **wind speed*** ...	Y such as, X1, X2
Measure	2	0	**0**	0.00	–	–	–
Characteristic	1	0	**0**	0.00	–	–	–
Overall	**71**	**59**	**53**	**0.84**			

Table 4. Press articles extracted relations validated by crisis expert concerning Ebola sanitarian crisis.

Concept	N	V	S	P	Examples of extracted instances		
					Instance	**Extracted data**	**Extraction rule**
Area	10	10	**7**	1.00	rural area	... *i headed to eastern sierra leone in some of the really **rural areas**.* ...	X = modifier Y
					metropolitan atlanta area	... *ambulance arrives in the **metropolitan atlanta area*** ...	X = modifier Y
					dallas area	... *into contact with mr. duncan attend four **dallas-area** public schools.* ...	X = modifier Y
People	13	10	**6**	0.73	the caregiver	... ***the caregivers** were often people who had survived smallpox themselves* ...	X is a Y
					sick people	... *find out who is harboring **sick people**, with potentially deadly consequences.* ...	X = modifier Y
					ms . sellu	... *the front line is stitched together by people like **ms. sellu**: doctors and nurses* ...	Y such as, X1, X2
Service	7	4	**4**	0.73	health service	... *charities and the united states public **health service** agreed to operate treatment* ...	X = modifier Y
					service uber	... *arranged through the online **service uber**, did not have direct contact* ...	X = modifier Y
					immigration service	... *lucy moreton, the head of the **immigration service** union,* ...	X = modifier Y
Measure	6	4	**2**	0.75	infection - control measure	... *when **infection-control measures** are poor, hospitals become "amplification points* ...	X = modifier Y
Threat	2	2	**2**	1.00	international threat	... *allowed the disease to mushroom from a local outbreak to an **international threat**.* ...	X = modifier Y
Resource	3	3	**1**	1.00	military medical resource	... ***military** and **medical resources** to combat the spread of the deadly virus* ...	X = modifier Y
					health resource	... *vice president of texas **health resources**,* ...	X = modifier Y
Risk	3	1	**0**	0.33	low risk	... *dr. frieden stressed that the passengers were a **low-risk** group.* ...	X = modifier Y
Opportunity	1	1	**0**	1.00	missed opportunity	... that **missed opportunity** has cost the lives of many people ...	X = modifier Y
Population	1	1	**0**	1.00	virus population	... dramatic change in the **virus population** could be explained by chance ...	X = modifier Y
Effect	1	0	**0**	0.00	–	–	–
Overall	**47**	**36**	**22**	**0.81**			

data related extraction. For each extraction, examples of extracted instances and the piece of text in which they appear are given.

Concept are ranked decreasingly regarding the number of extracted and surely validated instances (**S**). The validation task has been submitted both to a crisis management expert, having in mind the possible use of a crisis management Decision Support System and to another user mostly focused on the technical aspects of the extraction. Presented tables are built over the results of the domain expert validation. However, precision measures and similarity measure of the two validations can be found in table 9 for each type of crisis.

Consistency of precision measure and validation-induced bias

Regarding table 9, the first observation that we can make is that overall precision remains consistent over the different groups of documents. This observation applies for the type of crisis treated but also stands for the type of data source exploited, *press* or *academic*. The drop of precision in the case of Fukushima crisis, is essentially due to a high number of instances that are incompletely extracted (e.g. *nuclear plant* instead of *Fukushima Daiichi power plant*) and thus considered as not relevant, or not specific enough. Global consistency is encouraging as it reveals that the extraction rules and the extraction system are effective without regarding the type of crisis or the type of data source it is applied on.

Another observation can be made concerning the expert validation bias. Extraction results have been validated by two different experts adopting two different points of view, as explained in the previous section. Despite this

Table 5. Press articles extracted relations validated by crisis expert concerning hurricane Katrina crisis.

Concept	N	V	S	P	Examples of extracted instances: Instance	Extracted data	Extraction rule
Area	21	20	**16**	0.95	evacuation area	*... people living in the voluntary **evacuation area**, which includes most of metropolitan ...*	X = modifier Y
					new orleans area	*... of the dead collected so far in the **new orleans area**, more than a quarter of them ...*	X = modifier Y
					low-lying area	*... applied only to **low-lying areas** and not the city as a whole. ...*	X = modifier Y
Service	12	8	**6**	0.67	bus service	*... with a chicago-based **bus service**, the bus bank, to provide transportation ...*	X = modifier Y
					laundry service	*... for bills like 100-per-bag **laundry service**. ...*	X = modifier Y
					weather service	*... said timothy j. schott, a national **weather service** meteorologist. ...*	X = modifier Y
Threat	5	4	**4**	0.89	hurricane threat	*... were responding much as they had to many previous **hurricane threats**, ...*	X = modifier Y
					dire threat	*... him to raise the consciousness about the **dire threats**. ...*	X = modifier Y
People	4	4	**4**	1.00	stranded people	*... take to the **stranded people**, and to evacuate some on the buses, ...*	X = modifier Y
					old people	*... i looked down and there were about 15 or 20 **old people**. ...*	X = modifier Y
Event	5	3	**3**	0.67	fund-raising events	*... a plans to have **fund-raising events** when the preseason ...*	X = modifier Y
Effect	2	2	**2**	1.00	ecological effect	*... that contended the corps had failed to study **ecological effects**. ...*	X = modifier Y
					economic effect	*... reserve to blunt the **economic effects** ...*	X = modifier Y
Population	3	3	**1**	1.00	new orleans's population	*... while the bulk of **new orleans's population** evacuated before the storm, ...*	X = modifier Y
Good	1	1	**1**	1.00	dry good	*... building with water, **dry goods** and bath products ...*	Y such as, X1, X2
Resource	1	1	**1**	1.00	local state resource	*... **local** and **state resources** were so weakened ...*	X = modifier Y
Danger	1	1	**1**	1.00	potential danger	*... nursing home that faced the most **potential danger** from winds and flooding ...*	X = modifier Y
Strategy	1	1	**1**	1.00	white house strategy	*... crucial elements of a **white house strategy** to help mr. bush recover ...*	X = modifier Y
Risk	1	1	**0**	1.00	great risk	*... are often at the **greatest risk** of death and serious injury ...*	X = modifier Y
Factor	1	0	**0**	0.00	–	–	–
Overall	**58**	**49**	**40**	**0.84**			

Table 6. Academic articles extracted relations validated by crisis expert concerning Fukushima nuclear crisis.

Concept	N	V	S	P	Examples of extracted instances: Instance	Extracted data	Extraction rule
Event	51	37	**35**	0.82	flooding	*... against external events such as **earthquakes** and **flooding** ...*	Y such as, X1, X2
					earthquake		Y such as, X1, X2
					extreme wheather condition	*... external events – such as earthquakes, floods and **extreme weather conditions** ...*	Y such as, X1, X2
Organisation	39	31	**31**	0.86	vendor	*... organisations that support it – such as [...], **vendors** and their suppliers ...*	X = modifier Y
					nuclear safety organisation	*... of the fukushima accident, the **nuclear safety organisations** considered that provisions ...*	X = modifier Y
Accident	37	30	**27**	0.84	reactor accident	*... a database/matrix documenting the various **reactor accident** software programmes ...*	X = modifier Y
					nuclear accident	*... a **nuclear accident** (such as **an explosion** at a civilian nuclear power station) ...*	X = modifier Y
					an explosion		Y such as, X1, X2
Activity	44	22	**21**	0.64	nuclear regulation activity	*... **nuclear regulation activities** dealing with regulatory ...*	X = modifier Y
					nea activity	*... on priority areas for **nea activities** including ...*	X = modifier Y
Area	46	20	**19**	0.48	20-km radius area	*... an alert "contamination of people **20-km radius area**" is generated. ...*	X = modifier Y
					contaminated area	*... populations living in the **contaminated areas** face a chronic exposure ...*	X = modifier Y
					oceanic area	*... region of china and entire **oceanic area** around japan ...*	X = modifier Y
Factor	24	20	**18**	0.90	decontamination factor	*... been analysed (water balance, **decontamination factor** decreases ...*	X = modifier Y
					public opinion	*... so "**public opinion**" is key factor. ...*	X is a Y
					material factor	*... superposition of both **material factors** and ...*	X = modifier Y
Effect	13	11	**11**	0.92	thermal effect	*.. performed (e. g. the **thermal effects** created by two passive autocatalytic ...*	X = modifier Y
					human health effect	*... uncertainty in the predictions of dose and **human health effects** ...*	X = modifier Y
Risk	15	12	**10**	0.85	hydrogen risk	*... previous projects related to **hydrogen risk** ...*	X = modifier Y
					seismic risk	*... external hazards (**seismic risks**, flooding, ...*	X = modifier Y
					radiation risk	*... such direct **radiation risk** communication was quite helpful ...*	X = modifier Y
Resource	14	10	**9**	0.79	natural ressource	*... creative expression, and **natural resource** management (burke et al. ; 2006). ...*	X = modifier Y
Service	12	9	**8**	0.77	twitter	*... on popular services like **twitter**, automated programs ...*	Y such as, X1, X2
Strategy	9	9	**8**	1.00	mitigation strategies	*... implementing **mitigation strategies** during single and multi-unit events ...*	X = modifier Y
Objective	4	4	**4**	1.00	safety objective	*... wenra technical report is considered in the implementation of the **safety objectives** ...*	X = modifier Y
Sensor	5	3	**3**	0.75	photodiode sensor	*... generated a false signal in the **photodiode sensor** at a temperature ...*	X = modifier Y
People	3	3	**3**	1.00	injured people	*... **injured people** qualified as absolute emergency have a higher importance ...*	X = modifier Y
Population	4	2	**2**	0.57	ageing population	*... challenges of an **ageing population**. ...*	X = modifier Y
Characteristic	4	2	**2**	0.57	time characteristic	*... due to the **time characteristics** of the hazards ...*	X = modifier Y
Threat	2	2	**2**	1.00	external threat	*... defence-in-depth concept to ensure robustness against **external threats** ...*	X = modifier Y
Measure	12	1	**0**	0.04	[radiation levels] [...] disaster.	*... measure [...] radiation levels in their milieus since **the fukushima daiichi nuclear disaster**. ...*	Y such as, X1, X2
Danger	1	1	**0**	1.00	invisible danger [of radiation].	*... information to deal with the **invisible danger** of radiations ...*	X = modifier Y
Opportunity	1	0	**0**	0.00	–	–	–
Fact	1	0	**0**	0.00	–	–	–
Overall	**341**	**229**	**213**	**0.74**			

validation seems to lead to similar precision measures, the similarity computed between the two validated sets shows some divergences. Nevertheless, experts still seem to agree over a large majority of the instances, even if they adopt different views on validation criteria.

Influence of the type of crisis on the extraction

As stated, the influence of the chosen crisis on the precision of the extraction is not significant. However, if one focuses on the concepts for which instances are extracted, some differences emerge. Figure 1 represents the relative importance of the four most represented concepts (in terms of extracted and surely validated instances, aggregated over *academic* an *press* extractions) of each use case. It includes 9 concepts in total, as some concepts such as *Area* or *Event* are well represented in several use cases. This importance rate has been computed by dividing the number of surely validated instance for a given concept by the number of surely validated instances over all concepts.

Table 7. Academic articles extracted relations validated by crisis expert concerning Ebola sanitarian crisis.

Concept	N	V	S	P	Examples of extracted instances: Instance	Extracted data	Extraction rule
Service	76	57	**38**	0.79	vaccination	... barriers to accessing essential services, such as **vaccination** ...	Y such as, X1, X2
					all cancer treatment	... health services, such as **all cancer treatments** ...	Y such as, X1, X2
					schooling	... vital services, such as **schooling** ...	Y such as, X1, X2
Strategy	31	25	**21**	0.85	risk mitigation strategy	... with risk assessment and **risk mitigation strategies** in tandem ...	X = modifier Y
					rite strategy	... strategies relevant to this epidemic (e. g. , the **rite strategy** in liberia ...	X = modifier Y
					evidence-based strategies	... commitment to effective **evidence-based strategies**. ...	X = modifier Y
Area	29	20	**19**	0.75	western area	... high transmission in the **western areas** of both guinea and sierra leone ...	X = modifier Y
					metropolitan area	... widespread transmission in crowded **metropolitan areas**. ...	X = modifier Y
					densely populated area	... reduce cases in both **densely populated** urban **areas** ...	X = modifier Y
Activity	24	17	**11**	0.76	response work	... that **response work** is an activity that is frequently changing ...	X is a Y
					agriculture	... the threats to economic activities, such as **agriculture** ...	Y such as, X1, X2
Resource	15	10	**8**	0.78	u.s. resource	... decision to massively scale up **u. s. resources** ...	X = modifier Y
					financial resource	... allocate and audit **financial resources** according to rules ...	X = modifier Y
Organisation	8	8	**7**	1.00	civil society organisation	... governments, **civil society organisations** and international institutions ...	X = modifier Y
					the global fund [to fight aids]	... focused organisations, such as **the global fund** to fight aids ...	Y such as, X1, X2
People	7	7	**7**	1.00	infected people	... with some **infected people** staying in their communities ...	X = modifier Y
					vulnerable disadvantaged people	... the most **vulnerable** and **disadvantaged people** ...	X = modifier Y
Factor	14	12	**6**	0.82	sociodemographic factor	... societal infrastructure, **sociodemographic factors**, local unfamiliarity ...	X = modifier Y
					natural resources reserves	... revenues are influenced by many factors, such as [...] **natural resource reserves** ...	Y such as, X1, X2
Threat	10	7	**6**	0.72	infectious disease threat	... systems to detect and stop **infectious disease threats** ...	X = modifier Y
					high-profile threat	... infectious diseases are **high-profile threats** that alarm the world ...	X = modifier Y
Event	6	5	**5**	0.83	burial	... critical response events, such as case investigations and **burials** ...	Y such as, X1, X2
					key events	... outline dates in order to illustrate **key events** that triggered changes ...	X = modifier Y
Risk	12	8	**4**	0.67	security risk	... robust procedures and capacity for **security risk** assessments ...	X = modifier Y
					financial risk	... measurements for **financial risk** protection to ensure ...	X = modifier Y
					transmission risk	... responders about **transmission risks** and safety measures. ...	X = modifier Y
Objective	7	4	**4**	0.62	universal health coverage	... and we argue that **universal health coverage** is an affordable objective ...	X is a Y
Measure	12	5	**3**	0.42	prevention measures	... nurses and local organisations on **prevention measures**, ...	X = modifier Y
Characteristic	4	4	**3**	1.00	sudden change	... **sudden changes** of plans are a general characteristic of many types of disasters ...	X is a Y
Population	6	3	**2**	0.56	poorest marginalised population	... progress among their **poorest** and most **marginalised populations** ...	X = modifier Y
Effect	3	3	**1**	1.00	health effect	... the **health effects** of universal health care ...	X = modifier Y
Opportunity	3	3	**1**	1.00	time - limited opportunity	... ebola crisis in west africa presents a **time-limited opportunity** ...	X = modifier Y
Overall	**267**	**198**	**146**	**0.78**			

Table 8. Academic articles extracted relations validated by crisis expert concerning hurricane Katrina crisis.

Concept	N	V	S	P	Examples of extracted instances: Instance	Extracted data	Extraction rule
Resource	14	11	**9**	0.87	cleanup resource	... equitably assign debris **cleanup resources** to each region	X = modifier Y
					social resource	... find shelter near helpful **social resources**, while	X = modifier Y
					military resource	... many **military resources** begin arriving,	X = modifier Y
Area	12	11	**9**	0.95	disaster-affected area	... first-aid commodities to **disaster-affected areas** during the emergency response phase. ...	X = modifier Y
					widespread area	... solid waste are almost instantaneously deposited across a **widespread area**. ...	X = modifier Y
					canal street area	... major looting was generally limited to the **canal street area** ...	X = modifier Y
Event	10	8	**8**	0.84	the south eastern asian tsunami	... recent events such as [...], the **southeastern asian tsunami**, and **hurricane katrina** ...	Y such as, X1, X2
					hurricane katrina		Y such as, X1, X2
					flood event	... flow resulting from a **flood event** ...	X = modifier Y
Activity	9	9	**6**	1.00	mitigation activity	... showed its usefulness in evacuation planning and **mitigation activities**. ...	X = modifier Y
					staging activity	... found fema's pre-landfall **staging activities** to have been unprecedented in scale. ...	X = modifier Y
					recovery phase activity	... these **recovery phase activities** include cleaning up debris, providing temporary housing, ...	X = modifier Y
Effect	6	5	**5**	0.91	psychological effect	... the trying conditions of a disaster also have a **psychological effect**. ...	X = modifier Y
					macroeconomic budgetary effect	... **macroeconomic** and **budgetary effects** of hurricanes katrina and rita. ...	X = modifier Y
Service	7	7	**4**	1.00	weather service	... on monday, the national **weather service** offices, first responders ...	X = modifier Y
					mineral management service	... information from the u. s. **minerals management service** (mms) ...	X = modifier Y
Strategy	3	2	**2**	0.80	pre-positioning transportation strategy	... study different **pre-positioning** and **transportation strategies**. ...	X = modifier Y
Fact	3	2	**1**	0.75	official fact	... rumors immediately went from being unsubstantiated hearsay to **official fact**. ...	X = modifier Y
Objective	3	2	**1**	0.75	equity objective	... assigning resources to regions in order to achieve **equity objectives** ...	X = modifier Y
Population	3	2	**1**	0.75	local population	... that they did not warn the **local population** ...	X = modifier Y
Measure	4	1	**1**	0.25	performance measure	... system objectives, including **performance measures** ...	X = modifier Y
People	3	1	**1**	0.40	trained people	... with hundreds of **trained people** paying close attention to an emerging disaster ...	X = modifier Y
Risk	2	1	**1**	0.50	catastrophe risk	... private-sector company that specializes in **catastrophe risk** insurance ...	X = modifier Y
Actor	1	1	**1**	1.00	key actor	... none of the **key actors** managed to impose their frame on the general public. ...	X = modifier Y
Threat	1	1	**1**	1.00	potential threat	... evidence-based scenarios about a **potential threat** ...	X = modifier Y
Overall	**81**	**64**	**51**	**0.83**			

Table 9. Concepts of the crisis represented as classes in the ontology to instantiate (based on (Benaben et al. 2020) crisis metamodel) major concepts. Some *additional concepts* are used in order to cover more vocabulary.

Crisis	Fukushima			Ebola			Katrina		
	Press	Academic	All	Press	Academic	All	Press	Academic	All
Technical Overall precision	**0.72**	0.63	0.64	0.69	**0.74**	0.73	**0.84**	0.77	0.80
Domain Expert Overall precision	**0.83**	0.74	0.75	**0.81**	0.78	0.78	**0.86**	0.83	0.84
Validation similarity measure	0.58	**0.61**	0.60	0.59	**0.67**	0.66	**0.72**	0.62	0.66

An interesting fact is enlightened by this representation. The appearance of the *Accident* concept in the Fukushima crisis whereas it does not appear in other crisis reveals the difference in nature between Fukushima crisis, a man-made disaster, and Ebola and Katrina crisis. Similarly, it can be noticed that the concept *Service* is highly

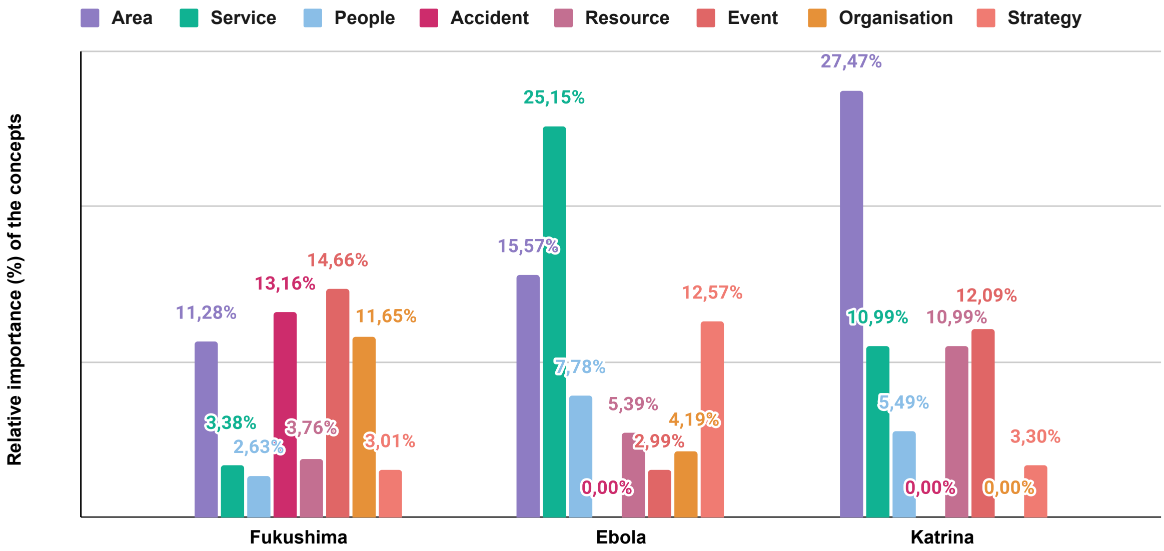

Figure 1. Representation of the relative importance of concepts having the most validated instances regarding the number of extracted instances (results aggregated over both press and academic extracted instances)

represented in Katrina and Ebola crisis probably revealing the strong impact of these crisis on different services related to health services instances for example. As well the concept *Area* is much more represented in Katrina crisis which could be explained by the importance that geographical areas took in Hurricane Katrina crisis (evacuation of populations, predictions of impacted areas, dimension of the risk area).

Rules disparity

Table 10 shows the precision results regarding the used extraction rules. τ is computed for each rule as the rate between the number of instances extracted by the rule and the global number of extracted instances. This table underlines the big disparities between rules as most of instances are extracted with *X = modifier Y* rule. Nevertheless, excepted for Fukushima case study, precision indicator over *X is a Y* and *Y such as, X1, X2* rules remains relatively high despite the low number of relations detected.

Table 10. Performances of the extraction and extracted instances distribution regarding extraction rules.

		Fukushima	Ebola	Katrina
X is a Y	τ	0.01	0.03	0.01
	P	0.56	**0.93**	0.5
Y such as, X1, X2	τ	0.10	0.12	0.03
	P	0.54	0.65	**0.86**
X = modifier Y	τ	**0.89**	**0.85**	**0.96**
	P	**0.78**	0.80	0.85

DISCUSSION AND PERSPECTIVES

The presented system and its applications on crisis management domain constitute a proof of concept of ongoing work. Applying the system to concrete use-cases emphasizes some further challenges to overcome and also new possible exploitation of extracted data. This section presents perspectives for the use of extraction results on the one side and directions that can be taken to improve such a generic system on the other side.

Using generic instances

Mostly, the extracted presented instances are related to the crisis treated. However, many terms are also extracted that represents instances that are not necessarily related to the studied crisis but to crisis in general. As these instances remain too generic (*concerned areas*, *urban areas*), they could not directly be used by a DSS in real-time and thus are not presented as examples.

However, this kind of instances can be valuable for at least two applications :

- First, they can be used to operate comparisons between crisis. Figure 2 shows diagrams in which each intersection contains common extracted instances for each pair of crisis. This overlapping rate is low as (1) the set of documents covered remain relatively low, (2) the chosen three crisis are voluntarily different and (3) the matching is computed through perfect match, ignoring instance describing the same object with different group of words. However, we hope that covering more data and improving matching measures allows to identify crisis that share similarities and to what extent (which concepts/instances). Making these comparisons and detecting similarities, between the ongoing Covid-19 crisis and previous sanitarian or economic crisis for instance, may support decision makers to benefit from past experiences and seize opportunities.
- As they still constitute an extension of knowledge by specifying generic concepts, they might be reused to extend the field covered by a crisis knowledge base. This use can be imagined in a second loop of extraction, where, without modifying the prior ontology, generic instances can be used as new concepts in order to detect more accurate instances (standing nearby in the text, for example).

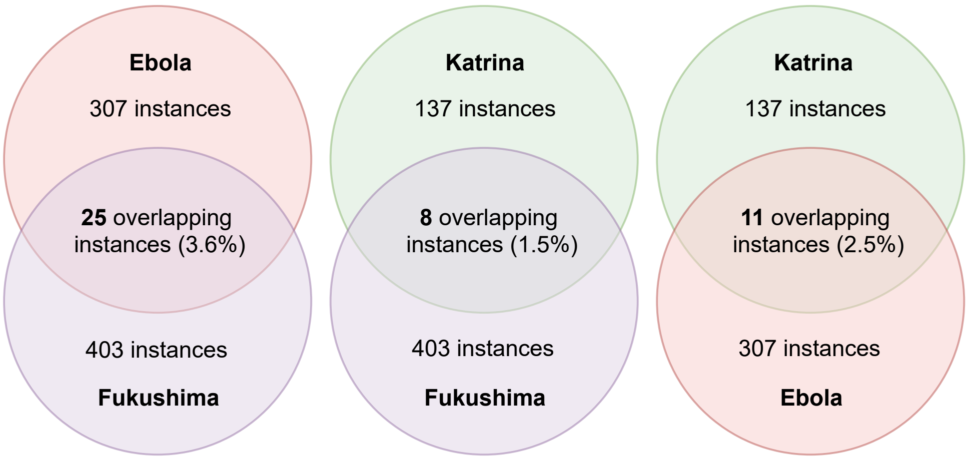

Figure 2. Overlapping instances of the different crisis (*academic* and *press* aggregated)

Concept dependency

The presented system relies essentially on the names of the concepts that are contained in the ontology. This approach guarantees ontology guided extraction. Nevertheless, given the polysemous nature of language, this approach presents some limits. Crisis applications presented here are a perfect example of this aspect because the terms used to define concepts are very common terms (*Event*, *Risk*, *Measure*), and have a very different signification in everyday life than in the context of a crisis. Hopefully, the hypothesis made that the explored documents are related to the domain of the ontology is a way to ensure a use of the vocabulary that matches the use made in the ontology. Still, exceptions can be found. The concepts *Measure* and *Area* for instance illustrate perfectly this default. In explored textual data related to crisis, the term measure is often used to define "measures" as taken actions whereas the concept *Measure* defines a quantified measure such as *radiation rate*, for instance. The concept *Area* in the crisis metamodel describes a delimited geographical area. Unfortunately, the term "area" tends to appear in *academic* data in order to define research areas which leads to instances that cannot be validated.

As a concept cannot be fully described with its name alone, ontologies classically provide a consensual definition of its concepts. Thus, using this definition to bring accurate information about concepts and filter possible errors could be a way to improve the precision of an extraction system based on the content of an ontology.

Benefit and limitations of automation of the ontology population process

The use of automated ontology population systems is a perspective that offers the possibility to save precious time when gathering information about a given crisis. Fully automated systems, because they do not require the intervention of a human being allow the processing of a large amount of data as explained in this paper. This constitutes one of the cornerstone towards effective crisis response. Besides the gathering of information to acquire knowledge during an event, knowledge bases can as well be populated using past experiences and feedback to support preparation. Nevertheless, some precautions must be adopted when applying such extraction systems as limitations exist.

First the use of automated methods, as long as they rely on a set of rules is precise. But still, because rules have a limited and predefined range of extraction, a huge part of the knowledge cannot be extracted. Thus, as soon as a part of the knowledge is extracted through these rules, experts might be needed in order to complete extraction finding undetected knowledge. Nevertheless, expert intervention can be driven by automated extraction as it reveals some interesting parts in the data, when many instances are detected in a given document for example.

As the presented approach is not dedicated to crisis management exclusively but to any kind of ontologies, the combination of automated and generic extraction method does not allow the extraction of the finest knowledge about a given case study with a generic crisis ontology. However, in order to extract more precise knowledge, more specific ontologies with fine-grained concepts can be used. These ontologies can then be processed with the same extraction framework as it can be adapted to any ontology. Still, the main objective of this paper is to apply the presented extraction system to a wide range of crisis. Hence, the study was not necessarily focused on the granularity of the used ontology.

CONCLUSION

In this paper, the use of a previously defined domain independent ontology population system has been illustrated on the specific case of crisis management. The extraction methodology has been applied with the same ontology classes onto three different types of crisis. The possible exploitation and application of the results of the driven extractions, based on simple extraction rules, is encouraging beyond the first objective of ontology population. Further work should be conducted on this field in order to (i) improve the developed extraction system using for example ontology properties linking classes in order to detect and link instances between each other and (ii) drive such automatically populated knowledge bases to be integrated in existing Decision Support Systems or simply give insight into ongoing crisis.

REFERENCES

Alobaidi, M., Malik, K. M., and Hussain, M. (2018). “Automated ontology generation framework powered by linked biomedical ontologies for disease-drug domain”. In: *Computer methods and programs in biomedicine* 165, pp. 117–128.

Altszyler, E., Sigman, M., Ribeiro, S., and Slezak, D. F. (2016). “Comparative study of LSA vs Word2vec embeddings in small corpora: a case study in dreams database”. In: *arXiv preprint arXiv:1610.01520*.

Barthe-Delanoë, A.-M., Truptil, S., and Bénaben, F. (2014). “Agility of crisis response: Gathering and analyzing data through an event-driven platform.” In: *ISCRAM*.

Bell, B. P. (2016). “Overview, control strategies, and lessons learned in the CDC response to the 2014–2016 Ebola epidemic”. In: *MMWR supplements* 65.

Benaben, F., Fertier, A., Montarnal, A., Mu, W., Jiang, Z., Truptil, S., Barthe-Delanoë, A.-M., Lauras, M., Mace-Ramete, G., Wang, T., et al. (2020). “An AI framework and a metamodel for collaborative situations: Application to crisis management contexts”. In: *Journal of Contingencies and Crisis Management* 28.3, pp. 291–306.

Bitencourt, K., Durão, F. A., Mendonça, M., and SANTANA, L. L. B. D. S. (2018). “An ontological model for fire emergency situations”. In: *IEICE TRANSACTIONS on Information and Systems* 101.1, pp. 108–115.

Boin, A., Brown, C., and Richardson, J. (2019). *Analysing a Mega-Disaster: Lessons from Hurricane Katrina*.

Buitelaar, P., Cimiano, P., Frank, A., Hartung, M., and Racioppa, S. (2008). “Ontology-based information extraction and integration from heterogeneous data sources”. In: *International Journal of Human-Computer Studies* 66.11, pp. 759–788.

Calcaterra, D., Cavallo, M., Di Modica, G., and Tomarchio, O. (2015). “A semantic framework to characterise and monitor territories affected by environmental risks”. In: *2015 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. IEEE, pp. 272–279.

Castillo, C. (2016). *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press.

Chasseray, Y., Barthe-Delanoë, A.-M., Négny, S., and Le Lann, J.-M. (2021). "A generic metamodel for data extraction and generic ontology population". In: *Journal of Information Science*, p. 0165551521989641.

Demchenko, Y., Grosso, P., De Laat, C., and Membrey, P. (2013). "Addressing big data issues in scientific data infrastructure". In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, pp. 48–55.

Endsley, M. R. (2001). "Designing for situation awareness in complex systems". In: *Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment*, pp. 1–14.

Ferrara, E., De Meo, P., Fiumara, G., and Baumgartner, R. (2014). "Web data extraction, applications and techniques: A survey". In: *Knowledge-based systems* 70, pp. 301–323.

Fertier, A., Montarnal, A., Barthe-Delanoë, A.-M., Truptil, S., and Bénaben, F. (2020). "Real-time data exploitation supported by model-and event-driven architecture to enhance situation awareness, application to crisis management". In: *Enterprise Information Systems* 14.6, pp. 769–796.

Fetter, G., Tech, V., Zobel, C. W., and Rakes, T. R. (2010). "A Multi-Stage Decision Model for Debris Disposal Operations". In: *The 7th International ISCRAM conference, Seattle, USA*, pp. 1–5.

French, S., Argyris, N., Smith, J. Q., Haywood, S., and Hort, M. C. (2017). "Uncertainty handling during nuclear accidents." In: *ISCRAM*.

Gaur, M., Shekarpour, S., Gyrard, A., and Sheth, A. (2019). "empathi: An ontology for emergency managing and planning about hazard crisis". In: *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*. IEEE, pp. 396–403.

Guan, Y., Shen, S., and Huang, H. (2015). "Assessment of the radiation doses to the public from the cesium in oceans after Fukushima Nuclear Accident." In: *ISCRAM*.

Hearst, M. A. (1992). "Automatic acquisition of hyponyms from large text corpora". In: *Coling 1992 volume 2: The 15th international conference on computational linguistics*.

Heng, L. and Tao, C. (2014). "Multiple attributes decision making method on social stability in nuclear accident scenario." In: *ISCRAM*.

Huynh, D. and Neptune, N. (2018). "Annotation automatique d'images: le cas de la déforestation". In: *Actes de la conférence Traitement Automatique de la Langue Naturelle, TALN*, p. 101.

Ishigaki, Y., Matsumoto, Y., Matsuno, Y., and Tanaka, K. (2015). "Participatory Radiation Information Monitoring with SNS after Fukushima." In: *ISCRAM*.

Jannach, D., Shchekotykhin, K., and Friedrich, G. (2009). "Automated ontology instantiation from tabular web sources—the AllRight system". In: *Journal of Web Semantics* 7.3, pp. 136–153.

Jones, K. S. (1972). "A statistical interpretation of term specificity and its application in retrieval". In: *Journal of documentation*.

Kaner, J. and Schaack, S. (2016). "Understanding Ebola: the 2014 epidemic". In: *Globalization and health* 12.1, pp. 1–7.

Karkaletsis, V., Valarakos, A., and Spyropoulos, C. (2006). "Populating ontologies in biomedicine and presenting their content using multilingual generation". In: *Programme Committee*, p. 51.

Kontopoulos, E., Mitzias, P., Riga, M., and Kompatsiaris, I. (2017). "A Domain-Agnostic Tool for Scalable Ontology Population and Enrichment from Diverse Linked Data Sources." In: *DAMDID/RCDL*, pp. 184–190.

Konys, A. (2019). "Knowledge repository of ontology learning tools from text". In: *Procedia Computer Science* 159, pp. 1614–1628.

Kopetz, H. (2011). *Real-time systems: design principles for distributed embedded applications*. Springer Science & Business Media.

Kunal, S., Saha, A., Varma, A., and Tiwari, V. (2018). "Textual dissection of live Twitter reviews using naive Bayes". In: *Procedia computer science* 132, pp. 307–313.

Landgren, J. (2015). "Insights from an ethnographic study of a foreign response team during the EBOLA Outbreak in Liberia." In: *ISCRAM*.

Liao, C.-h., Wu, Y.-f., and King, G.-h. (2019). "Research on learning owl ontology from relational database". In: *Journal of Physics: Conference Series*. Vol. 1176. 2. IOP Publishing, p. 022031.

Liu, X. and Gao, F. (2018). “An Approach for Learning Ontology from Relational Database”. In: *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 1–6.

Murphy, T. and Jennex, M. E. (2006). “Knowledge management systems developed for hurricane Katrina response”. In: *Third International Conference on Information Systems for Crisis Response and Management*.

OECD and Nuclear Energy Agency (2013). *The Fukushima Daiichi Nuclear Power Plant Accident*. OECD.

Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., Briant, J., Millet, P., Reinhard, F., Parkan, M., et al. (2016). “Combining human computing and machine learning to make sense of big (aerial) data for disaster response”. In: *Big data* 4.1, pp. 47–59.

Qin, Y., Guo, J., and Zhang, A. (2016). “A novel extreme learning fault diagnosis based supervision applied to mathematical formula contrastive analysis”. In: *Neurocomputing* 177, pp. 266–273.

Remolona, M. F. M., Conway, M. F., Balasubramanian, S., Fan, L., Feng, Z., Gu, T., Kim, H., Nirantar, P. M., Panda, S., Ranabothu, N. R., et al. (2017). “Hybrid ontology-learning materials engineering system for pharmaceutical products: Multi-label entity recognition and concept detection”. In: *Computers & Chemical Engineering* 107, pp. 49–60.

Roth-Berghofer, T., Adrian, B., and Dengel, A. (2010). “Case acquisition from text: Ontology-based information extraction with SCOOBIE for myCBR”. In: *International Conference on Case-Based Reasoning*. Springer, pp. 451–464.

Save the Children (2015). *The right to privacy in the digital age*. Save the Children.

Segault, A., Tajariol, F., and Roxin, I. (2015). “# geiger: Radiation Monitoring Twitter Bots for Nuclear Post-Accident Situations.” In: *ISCRAM*.

Skvortsov, N. A., Kalinichenko, L. A., and Kovalev, D. Y. (2016). “Conceptualization of methods and experiments in data intensive research domains”. In: *International Conference on Data Analytics and Management in Data Intensive Domains*. Springer, pp. 3–17.

Swaraj, K. and Manjula, D. (2016). “Fast Extraction of Article Titles from XML Based Large Bibliographic Datasets”. In: *Procedia Technology* 24, pp. 1263–1267.

Tekli, J., Charbel, N., and Chbeir, R. (2016). “Building semantic trees from XML documents”. In: *Journal of Web Semantics* 37, pp. 1–24.

Vicient, C., Sánchez, D., and Moreno, A. (2013). “An automatic approach for ontology-based feature extraction from heterogeneous textualresources”. In: *Engineering Applications of Artificial Intelligence* 26.3, pp. 1092–1106.

World Health Organization et al. (2015). *2015 WHO strategic response plan: West Africa Ebola outbreak*. World Health Organization.

Yeletaysi, S., Fiedrich, F., and Harrald, J. R. (2008). “A Framework for integrating GIS and Systems Simulation to analyze Operational Continuity of the Petroleum Supply Chain”. In: *5th International ISCRAM Conference, Washington, DC, USA*.

Zhang, Y. and Lu, Z. (2019). “Exploring semi-supervised variational autoencoders for biomedical relation extraction”. In: *Methods* 166, pp. 112–119.

Zschocke, T., León, J. C. V. de, and Beniest, J. (2010). “Enriching the description of learning resources on disaster risk reduction in the agricultural domain: an ontological approach”. In: *Research Conference on Metadata and Semantic Research*. Springer, pp. 320–330.