INFORMATION EXTRACTION FROM MESSAGES IN DISASTER MANAGEMENT

Stefan Werder, Christian Lucas, Hans-Peter Bähr
Institute of Photogrammetry and Remote Sensing (IPF), Universität Karlsruhe (TH), Englerstr. 7, 76128 Karlsruhe, Germany - (stefan.werder, christian.lucas, hans-peter.baehr)@ipf.uni-karlsruhe.de

ABSTRACT

Managing natural and man made disasters effectively requires detailed information about the current situation. As a consequence an up to date map containing both the status and location of incidents and resources in the field is needed. Usually this information is represented by a large quantity of incoming messages written in natural language text. This paper presents a comprehensive approach for extracting relevant information from these messages automatically. An ontology supported knowledge base for the requirements of disaster management has been developed based on a standard from the military domain. Concerning semantic gaps, the importance of context knowledge and temporal logic has been investigated. For a complete representation the reliability and uncertainties, inherent in each message, were considered.

Key Words: Disaster Management, Data Modelling, Information Extraction, Automation, Uncertainty

1 INTRODUCTION

Limiting the impact of damage and preventing further hazards are the main objectives during disaster response. These objectives can only be achieved efficiently, if the units working at the damage sites are coordinated and supported by one or more *emergency operation centres (EOC)*. The structure and size of these centres depend on the situation they have to cope with and the units they command. But all these centres have one important thing in common – the decisions they make can only be as good as the information available about the current situation. Therefore the provision of an up to date situation picture for all members of the emergency operation centre is of vital importance.

Most of the information addressed at the EOC is written down and the messages are forwarded to the members according to their responsibilities. Only two members have access to all information, one is the director of operations and the other one is responsible for the assessment, presentation and documentation of the current situation. The presentation includes the integration of relevant information into a map of the situation. The aim of the presented work is to contribute to the reduction of time needed for analyzing the messages and updating the map. In order to replicate the approach of a human operator with methods from information technology several steps are necessary. This paper focuses on the extraction, modelling and assessment of information from messages in disaster management including uncertain and fragmentary information. The different types of uncertainties that occur in the messages as well as a comparison of possible visualisation models were already discussed in (Werder et al. 2006).

1.1 Application of the SOKRATES system for disaster management

The prototype *SOKRATES* developed by the *Forschungsgesellschaft für Angewandte Naturwissenschaften* (FGAN, Research Establishment for Applied Science) has been designed to process messages about the position and movement of both friendly and hostile military units. The applicability of the system has been successfully tested with messages from a military scenario.

The processed military messages consist of two parts. The first part is a prologue which includes information about sender, position and time when the message was sent. The second part, which is much more difficult to handle, consists of a free form text of one ore more lines describing a situation. The *SOKRATES* workflow starts with the automatic transformation of relevant information from the messages into a formal structure, which is done by the information extraction component. Subsequently, the result from the information extraction is enriched by the usage of an ontology supported knowledge base. Finally the processed information about the units is stored in a database and displayed on a tactical map.

When the *SOKRATES* system and the demands for an up to date map for disaster management are compared side by side, the intersection between them is obviously. In both domains written reports containing free form text have to be processed in order to produce a map of the situation. The approaches used to attain this goal are to a high degree similar. Despite these similarities there are also differences, mainly based on the differences between the two domains. Due to the higher regularisation in the military domain the messages are formulated in a more consistent manner. Another important difference is the information that is displayed in the map. The *SOKRATES* system focuses on the detailed position and movement of units, whereas during disaster response the position and detailed status of damages and hazards is of primary importance. The granularity of the position of units during disaster response is often only needed at the scale of damage sites and standby areas. Additionally the fact that a unit is on the move from one of them to another is of importance.

The significant benefit from adapting the *SOKRATES* system for disaster management is that an existing system can be modified rather than designing and implementing a new system from scratch. In the following, some important aspects of this adaptation are presented, including the information extraction system (chapter 2) and in more detail the modifications of the knowledge base (chapter 3). Another important component that has not yet been considered thoroughly in the prototype is the modelling and use of context knowledge (chapter 4). Context comes also in handy when dealing with uncertainty and reliability (chapter 5), two important facets of the dynamics of man made or natural disasters. Aspects of temporal logic (chapter 6) play also an important role, because managing disasters is always a time critical task.

2 INFORMATION EXTRACTION SYSTEM

Systems that perform information extraction have to be able to “find and link relevant information while ignoring extraneous and irrelevant information” (Cowie and Lehnert 1996, p. 81). Based on this definition, in *information extraction (IE)* given texts are not examined by a full syntactic analysis of every sentence, because this approach is too slow and error-prone. Instead a shallow parsing strategy is used to extract only relevant information fragments. Normally information extraction systems are tailored for a specific task and domain. This compromise is made for obtaining high processing speed and accuracy. The domain dependence is realised by a domain-specific lexicon and extraction rules. In their *Introduction to Information Extraction Technology* Appelt and Israel (1999) identify the four primary modules of every IE system. These build a process chain, which begins with tokenization, followed by morphological and lexical processing, syntactic analysis and finally domain analysis. Another facet of processing natural language texts is that IE systems are normally limited to text written in one or few languages.

The processed reports in *SOKRATES* as well as the messages from disaster management used in this project are written in German. Therefore the information extraction core system *Saarbrücker Message Extraction System (SMES)* has been used in the *SOKRATES* system and is also used in its adaptation for disaster management. The architecture as well as application scenarios of *SMES* are described in detail in (Neumann et al. 1997). Additionally the adaptability of *SMES* for other domains as well as its performance has been demonstrated by the information extraction from texts about a soccer championship (Neumann and Declerck 2001).

In the *SOKRATES* system the *SMES* has been primarily modified and enlarged regarding the two components lexicon and transducers. Thereby several unit types and location names have been added to the lexicon. The transducers are actually finite state transducers, which are finite state machines with two tapes. They translate content of their input tape to their output tape. The output from one transducer can serve as the input for another transducer. Information hidden in the written text is extracted into *typed feature structures*, which is a formalism widely used in computational linguistics (cf. (Pollard and Sag 1994)). Examples for the future structures used in *SOKRATES* are given by Hecking (2004).

For the adaptation of *SMES* for disaster management the same types of modifications apply. Information about the unit types, for example the proper names of all vehicles of fire brigades, have to be added to the lexicon. In order to be able to classify locations, the names of all streets, towns and important points of interest must also be integrated into the lexicon. Additionally the lexicon has to be revised to hold the terms that are used to describe the damage states, for example the states of a fire. The second part of the modifications concerns the transducers, which have to be rewritten to extract both the movement of units and information about damage locations and states. As already stated, the granularity of the position of units doesn’t need to be very high. Because the modifications are specific to German language and to the *SMES* system, this paper focuses on the modelling of the ontology. By defining the information that has to be extracted, the specifications for the IE system follow implicitly, for example the form and content of feature structures.

3 MODELLING THE ONTOLOGY

The extracted information from each message adds additional knowledge to the situation picture. In order to represent this knowledge formally an ontology is used. According to Gruber (1993) the term *ontology* is defined as “an explicit specification of a conceptualization”. Because it is impossible to specify knowledge completely, an ontology is always restricted to a set of objects that it is able to represent, the so-called *universe of discourse*. These objects are defined in an ontology by classes and the relationships between them. A human-readable text description of both, along with rules that constrain interpretation and usage of objects, finally add meaning to the ontology.

An ontology can also be seen as a *semantic net*, in which nodes are represented by classes and edges by relationships between the classes. Because an ontology is generally spoken a data model, it can also be modelled and stored in several other ways. This includes, but is not limited to, databases, files in the Extensible Markup Language (XML) format or platforms like *Protégé*. The latter is also a good example for the diverse applications of ontologies, because *Protégé* has been developed by Stanford Medical Informatics (2007) for the field of biomedicine and is now also in use for intelligence gathering, corporate modelling and more.

In the following the usage of ontologies in two other domains – the military and the disaster management domain – are presented in more detail.

3.1 The Command and Control Information Exchange Data Model (C2IEDM)

The Multilateral Interoperability Programme (MIP) is an association of 24 nations and several organisations, such as military agencies. It aims to achieve international interoperability of Command and Control Information Systems, which is an important demand of modern armies. Nowadays they are faced with multinational, combined and joint operations, which have to be coordinated efficiently. These operations include conventional general war, asymmetric conflict and terrorism, but also crises response actions. The intersection between disaster management and military operations can be seen here again, because the word "crises" refers also to natural and man made disasters.

The core of the MIP solution is the *Command and Control Information Exchange Data Model* (*C2IEDM* 2005), which models the information that needs to be exchanged during joint operations. According to this definition, the *C2IEDM* has not been designed with the goal to model and store all the information that would be normally required by a national system. Keeping that difference in mind, the *C2IEDM* can be seen as an ontology for information exchange, because the minimum requirement demanded by the MIP is that the "meaning and relationships of the information to be exchanged" (C2IEDM 2005) need to be preserved.

The first version of the *C2IEDM* has been released at the end of 2003 and the actual version 6.15e is from October 2005. The successor will be the *Joint C3 Information Exchange Data Model* (*JC3IEDM* 2006) which is already under construction and shall be released in 2008. It incorporates minor improvements and additional data from NATO standards. Nevertheless, as *C2IEDM* is the actual data model and as the *SOKRATES* ontology as well as the ontology for the disaster management are based on it, the basic concepts of this standard will be presented in the following.

The *C2IEDM* is defined by three different views on the data model. The first is the Conceptual Data Model which provides a high level view of the information and therefore represents concepts only in a generalised way. The second is the Logical Data Model which goes into detail with the provision of an entity-attribute-relationship diagram. The third one is the Physical Data Model which defines the structure of a *C2IEDM* compliant database. As an alternative a XML-Schema is also provided by the MIP along with the other documents.

From the 194 entities defined in the Logical Data Model only 15 are independent. Independent means, that the identification of these entities does not depend on any other entity. These independent entities (see Figure 1) provide an overview over the data model, which is even more generalised than the high level view mentioned above.

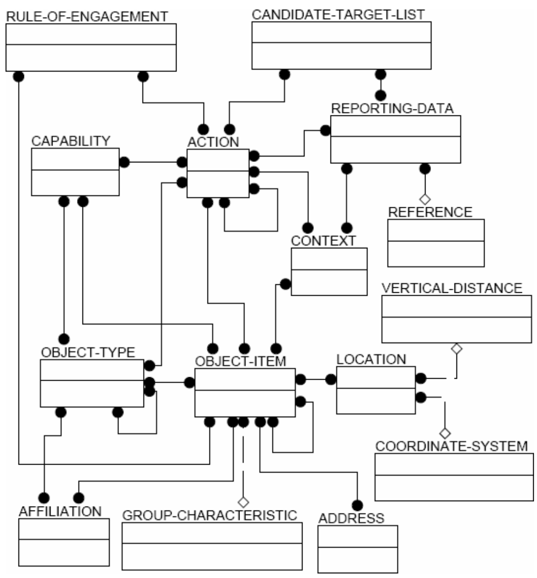

Figure 1: Independent entities of the C2IEDM (C2IEDM 2005, p. 24)

The *universe of discourse* of the *C2IEDM* is defined by the two entities *Object-Type* and *Object-Item*. A particular object in the *C2IEDM* is modelled with both entities. First by the attributes which values are common among all objects of that type (e. g. track width of a vehicle), as defined in *Object-Type*. Second by the attributes which values can differ between all objects of a type (e. g. call sign or actual fuel level of a vehicle), as defined in *Object-Item*. Because *Object-Item* is an instance of *Object-Type*, all common attributes are inherited by the individual items. The distinction of the two entities implicitly divides the attributes of an object into the more static ones relating to *Object-Type* and the more dynamic ones relating to *Object-Item*. Although this "division rule" helps sometimes understanding the data model better, it is not universally valid. A person's blood type, for example, can differ between different persons but is obviously not a dynamic value. At the top level of the hierarchies of *Object-Type* and *Object-Item* five different subtypes of objects, which are shown in Table 1 (based on (C2IEDM 2005)), can be modelled in the data model.

Table 1	
Name	*Description*
Facility	Objects that are built, installed or established to serve some particular purpose (e. g. airfield, bridge, road)
Feature	Objects that encompass meteorological, geographic and control features of military significance
Materiel	Objects that are equipment, apparatus or supplies of military interest; further subdivided into consumable material type (e. g. ammunition, food, fuel) and equipment type (e. g. aircraft, vehicle, weapon)
Organisation	Objects that represent administrative or functional structures (e. g. civilian, government including military organisation)
Person	Human beings

An example for relationships between entities in the *C2IEDM* is the entity *Location*. It is used to specify the geometry of objects, for example the shape of a rescue corridor. It is also used for the placement of objects or their geometry in relation to the Earth's surface or to other objects. A *Location* can be a one of the types point, line, surface or a geometric volume. The entity *Address*, shown in Figure 1, is not related to *Location* in the data model, because it holds only information about how a destination can be accessed in the context of communication. An *Address* is either an electronic address, which can be used via a network service, or a physical address, which can be reached for example via postal services.

The entity *Action* is used for representing activity in the *C2IEDM* and is divided into two subclasses. The first one includes planned or carried out activities as part of military operations (*Action-Task*), which use objects both as resources and objectives. *Action-Task* covers a wide spectrum of activities, for example ambush, block, confiscate, construct, identify, move and provision of shelter. The second subclass includes activities which occur or have occurred (*Action-Event*), but for which in contrary to Action-Task the plan is unknown. Examples for *Action-Event* are civil demonstrations, escaping, friendly fire and terrorism. But also disasters are covered by *Action-Event*, as shown later.

Reports by persons or organisations change the situation picture. Therefore the information about source, quality and timing is captured by the entity *Reporting-Data*. The information about the "what" is stored in numerous entities in the *C2IEDM*, for which the *Reporting-Data* provides applicable information. In case the information is derived from external data, for example from electronic mail, the entity *Resource* can be used to hold one or more *Reporting-Data* instances that correspond to the information pieces contained in this external data.

In the data model the entity *Context* can be used to group data in order to bundle information already available. For example a group of data can consist of some reports, which are evaluated by an intelligence officer to create a new report. *Context* can also be used to specify the prerequisites and estimated results of an *Action*. Furthermore it can be used to hold all information about a situation in the past or in the future.

The other independent entities of the *C2IEDM*, as shown in Figure 1, are less important with regard to the modelling of the Disaster Management Data Model and are therefore not discussed here.

3.2 The SOKRATES ontology

The ontology of the *SOKRATES* system is modelled in *Protégé*. Its object hierarchy is derived from the *C2IEDM*, whereas some concepts were simplified and others enhanced. The ontology is primarily used to integrate semantic and pragmatic knowledge (Schade and Frey 2004) into *SOKRATES*.

When building an information system not only the information exchange has to be considered, but also the information storage. Using the concepts of ontologies, this storage can serve as a knowledge base. The analysis of the applicability of the *C2IEDM* for that purpose leads to two important considerations. First, the *C2IEDM* has been designed for the purpose of information exchange, and not for storage of all relevant information that a stand alone system normally needs. Second, the Physical Data Model defines the structure of a *C2IEDM* compliant database, which is in fact a relational database. The result from the first consideration is the considerations that additional concepts and classes may have to be introduced. These extensions can be guided by the ones from the *C2IEDM* and other standards. The second aspect has to be considered with regard to the practical implementation of that information system. Modern programming languages are normally object-oriented. In these languages classes use inheritance rather than relations. The differences between *object-oriented programming (OOP)* and relational databases result in conceptual and technical difficulties, which are known as the Object-relational (impedance) mismatch. Some of these difficulties can be solved by Object-Relational Mapping (ORM) tools, but often some drawbacks remain unsolved.

An object-oriented approach for the ontology avoids these difficulties. Therefore the needed classes and concepts of the *C2IEDM* ontology have to be modelled in terms of inheritance. The most obvious change is the unification of the two entities *Object-Type* and *Object-Item* into a single entity named *Object*. This has been carried out, because the distinction between attributes, which values are common, and attributes, which values differ between all instances of a particular object type, is not drawn in OOP. Nevertheless, the common values can be set in the default constructor of an object. When an object-oriented approach is used, the position where the information is stored also changes, which will be shown by an example. In the *C2IEDM* the information about a bridge is stored in a total of six entities (number of attributes in columns) – *Object-Item* (3), *Facility* (8), *Bridge* (4), *Object-Type* (4), *Facility-Type* (2) and *Bridge-Type* (2). With the concept of inheritance and abstract classes the same information about a bridge can be stored in a single class *Bridge* (13), which is a child of the abstract class *Facility*, which is on his part a child of the abstract class *Object*. Summed up, the number of attributes for the relational approach is a total of 23. In contrast the OOP approach is able to store the same information with 10 attributes less, because foreign-keys and discriminators are not needed. Additionally, the information is stored directly within the object in the OOP approach and is not distributed among several objects.

Protégé allows the usage of inheritance and abstract classes; hence it has been used for modelling the knowledge base. The information inside *Protégé* then can be accessed within the implementation code of the information system by method calls. The object-oriented approach is also used by the Disaster Management Data Model, which is based on the *SOKRATES* ontology.

3.3 The Disaster Management Data Model (DM²)

The *Disaster Management Data Model (DM²)* has been developed with focus on the adaptation of the SOKRATES system for an application in disaster management. It does not to cover all facets, for example resource management is not considered in detail yet. This aspect is already incorporated rudimentary in the *C2IEDM* in *Consumable-Material-Type*, which denotes material that has to be reordered if it runs short (cf. Table 1). However, resource management in disaster management is more complex, but is already covered by initiatives like the Emergency Data Exchange Language (OASIS-EDXL 2005). Although the development of the *DM²* is motivated by the integration in the *SOKRATES* system and based on the *C2IEDM*, many considerations relating to the data model apply to ontologies in the domain of disaster management in general.

The most important entities of the *DM²* are shown in Figure 2. In the following the differences to the *C2IEDM* (cf. Figure 1) and the new concepts which were introduced are discussed.

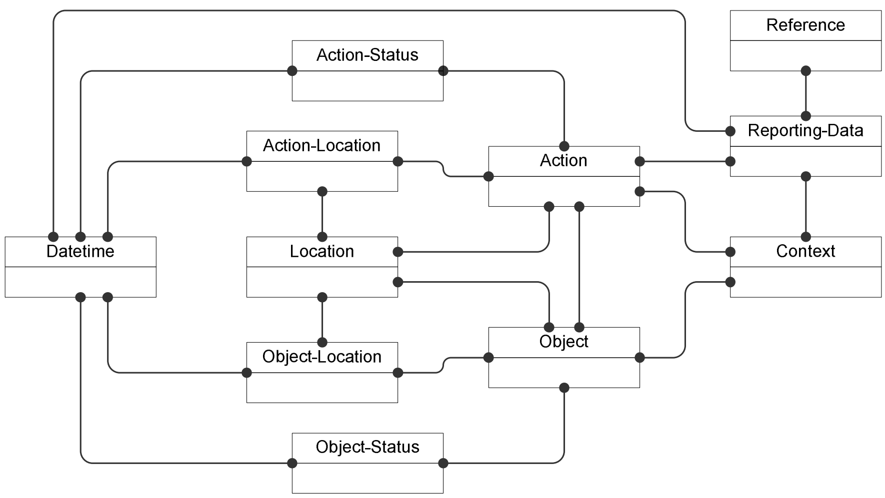

Figure 2: Important entities of the Disaster Management Data Model (DM²)

As a consequence from the object-oriented approach for the *DM²* ontology, the entity Object contains the whole *universe of discourse*. The five different subtypes of objects were kept, but in order to adapt to the disaster management jargon the military term “materiel” has been renamed to “resource”. The real differences are however in the instances of the objects, which are summarized in Table 2 below.

Table 2	
Name	*Description of the changes*
Facility	In the C2IEDM surprisingly no buildings were modelled, but of course they play a significant role in disaster management. Therefore the concept of buildings was added, including their role (e. g. school, hospital, sports field) which is important for many decisions, like evacuation. Additionally places were modelled, because of their common usage as gathering places.
Feature	The concept of control features was adapted for the DM² by adding for example damage sites, operation sections and standby areas.
Resource	Although the C2IEDM also offers for example a vehicle type "Firefighting", the actual fleet of vehicles of fire brigades is much more complex. The resource hierarchy has to be modelled according to the actual conditions and is therefore very specific.
Organisation	In the DM² based on the Resources new unit types were added (e. g. a fire unit consists of several vehicles) and also the different tactical levels (e. g. the emergency operation centre (EOC))
Person	Not changed, but in analogy to "Person-Language-Skill" in the C2IEDM the profession of a person along with individual skills can be added

In most cases disaster management takes place in (densely) inhabited areas. The address of a building is important geo-related information, and has to be stored inside the knowledge base along with its coordinates. Additionally a geo-coder must be provided to connect them.

The entity *Action* has also been modified for the *DM²*. The planned activities in the subclass *Action-Task* have been extended to the tasks of disaster management, like fire fighting. Although the *C2IEDM* provides the classification of a particular *Action-Event* as a disaster (e. g. earthquake, flood, fire, volcanic eruption), this approach provides not enough detail for disaster response. Normally several additional attributes are needed for a each type of disaster, for example in case of an earthquake the depth and magnitude are important, whereas for a building collapse the damage type has an significant influence on the decision making process.

Written report forms used in disaster management often provide a field to indicate the priority of the message. The Common Alerting Protocol (OASIS-CAP 2005) provides an element called *urgency* with the possible values "immediate", "expected", "future", "past" and "unknown", which denote the available response time. In the *DM²* the scale defined by the German *Bundesamt für Bevölkerungsschutz und Katastrophenhilfe* (literal: Federal Office for Population Protection and Disaster Relief) has been used. In the DV 810 (1977) the four priority levels "Einfach" (lit. routine), "Sofort" (immediate), "Blitz" (flash) and "Staatsnot" (lit. state emergency) are defined. Although in the *C2IEDM* already information about the quality of *Reporting-Data* is defined, for disaster management a more detailed consideration of vagueness is needed, which will be shown in chapter 5.

It has already been stated, that disaster management is a time critical task. Therefore in the *DM²* the entity *Datetime* has been introduced to take this important fact into account. Although time is already incorporated in the *C2IEDM* in several entities including *Action*, *Context* and *Reporting-Data*, for a refined view of the course of time several aspects have to be considered, which are presented in chapter 6.

Because of their importance for the *DM²* the entities *Action-Location*, *Action-Status*, *Object-Location* and *Object-Status* along with their relationships are part of Figure 2. These entities capture the highly dynamic situation in terms of time (*Datetime*), shape and position (*Location*) and changing attribute values (*Status*). The principles behind this approach can be best shown – without loss of generality – by the example of the two objects *Building* and *Person*. A person is able to move around, so tuples with (*Person*, *Datetime*, *Location*) track the position in *Object-Location*. Because the health status of a person can change, the value of the attribute has to be tracked in tuples (*Person*, *Datetime*, *Attributes describing the status*) in *Object-Status*. In contrast to the "division rule" between *Object-Type* and *Object-Item* of the *C2IEDM*, in the *DM²* the entity *Object-Status* strictly holds only values that can change. The considerations about *Object-Status* also apply to the example of a building. Here the damage and fire state are stored along with their timestamp. But as buildings are unlikely to change their position, they are directly linked to the entity *Location*.

Context in the *Disaster Management Data Model* is more than just labelling a group of data. It is used as an important part of the knowledge base as discussed in the following chapter.

4 THE ROLE OF CONTEXT

A substantial decision base for the emergency operation center is the situation map, which refers to reports from different damage sites (cf. chapter 1). The on-site units concentrate all impressions and individual observations in reports, which represent a generalized picture of the situation. The abstraction level of these messages depends on the source and the type of observation. Often it is impossible to give precise or complete predications, for example about the number and health state of injured persons. In addition fuzziness is inherent in messages, because of the free text form. These aspects contain a high degree of uncertainty and unreliability, which are discussed in detail in chapter 5.

In some cases the messages include implicit given information. For example from a report about smoke an incipient fire can be inferred. A human operator is able to understand the full information content of such messages, which is not completely revealed by the text. So-called *semantic gaps* are a conceptual summary of all these aspects.

With the aid of semantic considerations and heuristic acceptances, a human operator is able to handle all aspects of the reports, based on the knowledge about the coherences. The natural approach to the enrichment of reports has to be adapted for the automation. This is made possible by the inclusion of basic knowledge, like interdependence of smoke and fire, and context knowledge, like the interaction of different events.

The basics for using context knowledge in this way have to be provided by the knowledge base. In the *C2IEDM* the context is defined as a collection of information which provides the circumstances, conditions, environment, or perspective for a situation (cf. chapter 3). In the ontology of the *DM²*, context is a powerful instrument for semantic enrichment. The implementation is related to the *C2IEDM*, by the cross linking of the main entities *Action*, *Object* and *Reporting-Data* (cf. Figure 2). The philosophy behind the context differs nevertheless, because the context in the *DM²* is built in task-related hierarchies. Hence the meaning of this is to bundle information at the base. The main context of a strong damage event is subdivided in sub-contexts which are subdivided once more, until the smallest organization device is reached. According to that, the ontology is able to identify all coherences for semantic enrichment.

5 UNCERTAINTY AND RELIABLITY

The different facets of the nature of reports written in natural language, like semantic gaps or the sources of vagueness, were already discussed in chapter 4. The usage of context knowledge was presented as a powerful instrument for the semantic enrichment. Nevertheless, in order to completely represent the message content in the situation map, it is necessary to evaluate the vague and fuzzy facts. These interdependent values of uncertainty and reliability are basically present in all reports. For the knowledge acquirement vague statements are not adverse. So Russell (1923) warned against believing that vague knowledge must be wrong. On the contrary, it is more likely that a vague statement fits the truth, than a precise one.

Vagueness (here in terms of uncertainty, reliability, fuzziness etc.) is a specific characteristic of the reports, which results from the verbally representation in free text form (cf. chapter 4). The operational standard for the German disaster management domain, the *DV 100* (1999), explicitly emphasizes that situation observation can be incomplete, inaccurate, contradictory and sometimes even wrong. To handle and model this vagueness, a separation of the variable types is advisable.

Within the entity *Reporting-Data* of the *C2IEDM* the attributes *accuracy*, *reliability* and *credibility* are differentiated as shown in Table 3 (C2IEDM 2005, p. D-39). The attributes are treated as independent parameters with the possibility of a discrete appraisal by the allowed values.

Table 3

Attribute	*Definition*	*Values*
accuracy	The specific value that represents, for intelligence purpose, the general appraisal of the subject matter in graded terms to indicate the extent or degree to which it has been judged to be free from mistake or error or to conform to truth or some recognized standard value.	confirmed probable possible doubtful improbable truth cannot be judged
credibility	The specific value that represents, for normal operational use, the degree of trustworthiness of the data referenced by a specific Reporting-Data.	indeterminate reported as a fact reported as plausible reported as uncertain
reliability	The specific value that represents, for intelligence purpose, the general appraisal of the source in graded terms to indicate the extent to which it has been proven it can be counted on or trusted in to do as expected.	completely reliable usually reliable fairly reliable not usually reliable unreliable reliability cannot be judged

The philosophy of categorizing vagueness in the *C2IEDM* differs from the *DM²*, which distinguishes between *uncertainty* and *reliability*. The attribute *uncertainty* contains all vagueness which relates to the characterization of a reported fact. This includes fuzziness in statements for descriptive features like dimension, location or quantity. *Reliability* relates to the reported fact of the statement. This is primary affected by the source of the message.

An example can be shown by the phrase “probably 7 injured people”. The fragment of “probably 7” is a characterization and represents the *uncertainty* in the sentence. Fuzzy here is the meaning of 6 to 8. The reported fact is the fragment “injured people”. The *reliability* of this fact is influenced by the source of the message and the type of observation. This can be derived from the *DV 100* (1999), which gives in addition instructions for the handling of reports. According to the *DV 100*, the reliability of a staff member in a search and rescue unit will be assumed higher than the reliability of a passerby. Additionally the type of observation is differentiated between statements of a third party, based on own observations or assumptions of the author.

The attributes of vagueness within the *DM²* are interdependent. If a characterization, like “probably 7” is given, a higher *reliability* of the reported fact is generally assumed. This based on the deliberation that specifications support the reported fact, because detailed knowledge is available.

The assessment of the *uncertainty* and *reliability* within the *DM²* is generally based on confidence factors in the range of $[0, 1] \in \mathbb{R}$, with the meanings $[0 \equiv \text{no trust}]$ and $[1 \equiv \text{full trust}]$. The factors are informal measures of the trustfulness, which represent the degree of truth and not the probability of a thesis. The usage of these confidence factors allow combinations of influence factors to enhance the assessment, based on the methods of possibilistic and probabilistic reasoning. The result is a combination of *uncertainty* and *reliability* to represent the vagueness of a statement in a general value.

The source of the confidence factor is case-related quite different. For example estimations for fuzziness of the adverb phrase “probable” (given by the example above) are covered by a variety of studies, like Kipper and Jameson (1994) or Renooij and Witteman (1999). These studies declare the probability of the expression “probable” to 85% which conforms to a confidence factor of $[x = 0.85]$. Following this approach, every fragment of a statement gets an empiric parameter, which represents the vagueness.

6 TEMPORAL LOGIC IN THE DOMAIN OF DISASTER MANAGEMENT

Temporal reference is an essential feature in the domain of disaster management. All messages receive a lot of time stamps without considering the sender or receiver. The report forms allow documenting the course of time in five different ways. It is listed when the sender observes the scenario, when the message was sent, when the message arrives, etc. Additionally, the time or temporal relations are frequent contents of the messages, for example how much time is needed to evacuate a building or to extinguish a fire. This seems redundant, but the documentation of the timeline allows the detection of gaps or weak points in the so-called *report chain*. The actuality of a report or content can also be evaluated.

A characteristic feature of the disaster management domain is to report the *estimated time of arrival (ETA).* The emergency operation centre assumes the unit arrives, unless a contradictory message is available. Constraints like this are representable by the so-called Allen relations which define the thirteen basic relations of interval algebra (Allen and Ferguson 1994) shown in Figure 3.

Relation	Symbol	Inverse Symbol
X *before* Y	<	>
X *meets* Y	m	mi
X *overlaps* Y	o	oi
X *during* Y	d	di
X *starts* Y	s	si
X *finishes* Y	f	fi
X *equal* Y	=	=

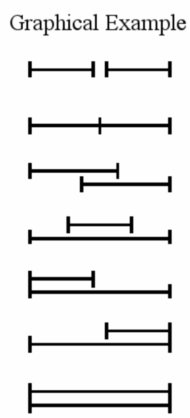

Figure 3: The basic relations of interval algebra (Puppe 1991, p. 70)

These interval relations allow the description of possible temporal links between different actions in a way of relative timings. An example is the relation between the end of a fire (*Y*) and the end of extinguishing a fire (*X*) which are linked by the relation *X finishes Y*, or the inversed relation *Y finished by X*.

In order to characterize absolute timed actions, the temporal model of Allen has to be enlarged by a temporal model of moments which is described by Gerevini and Schuber (1994). Within this model of time the interdependence between moments consist of the basic logical relations <, >, ≤, ≥, = and ≠. It is also necessary to dissolve the relative duration *X* to a starting point X^- and an endpoint X^+. Considering these facts, it is possible to express the relation *X before Y* according to Eq. 1.

$$X \text{ before } Y \quad (X^- < X^+) \wedge (Y^- < Y^+) \wedge (X^+ < Y^-)$$

Equation 1

In order to represent the requirements of disaster management, it is necessary to model reports like "we shall arrive in 10 minutes". This message includes the temporal unit "minute", which is essential to implement a basic time unit for describing durations. The temporal information of this simple message can also be displayed in the logical view $(X^- +10 = X^+)$ with the meanings of *[*$X^- \equiv$ *now], [+10=* ≡ *relation in min]* and *[*$X^+ \equiv$ *arrival]*. It is also possible to represent complex timings unambiguously, like the interaction of many devices or different tasks. This is shown in Figure 4, exemplary for the action of a search and rescue unit. The search for survivors is defined by *X*, eliminating a blockade by *Y* and the rescue of a person by *Z*. The relationships – *X overlaps Y*, *Y overlaps Z* and *X meets Z* – are used to express the situation.

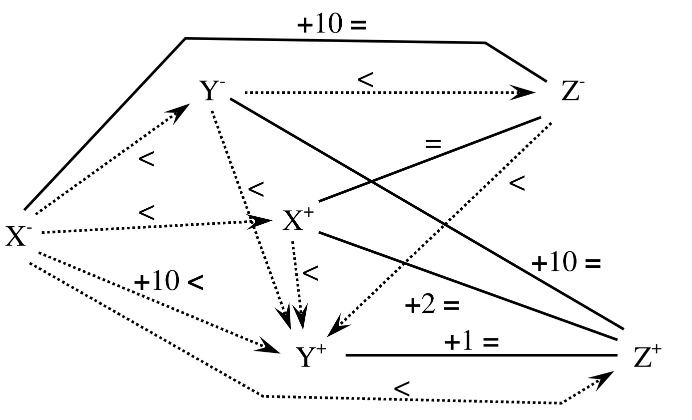

Figure 4: View of the absolute (solid lines) and relative (dashed lines) temporal relations (in min)

Figure 4 associates the absolute and the relative temporal relations. The solid lines represent the minimal configuration to describe the whole situation unambiguously and absolutely. The information needed can be extracted from the report forms. The dashed lines represent the synoptic view of possible temporal relations which are derivable from the given information. This functionality allows to test the logical consistence of the temporal information or to define temporal conditions. To fulfill that purpose redundant information is needed, which is given by the quantity of messages.

The temporal logic in the domain of disaster management as explained in this chapter, is modeled by a *semantic net*. The points of time, like X^-, are the nodes and the relations, like < or "overlaps", are the edges. The ontology is based on the knowledge base of the *DM²* and provides an efficient way to handle these.

CONCLUSION

Extracting information from messages in disaster management is an extensive challenge. Some of the arising problems have been already solved in the military domain, so an adaptation of their expertise is promising. For the application in disaster management adjustments of the *SOKRATES* system are necessary. Modifications of the information extraction component mainly arise from different terms and concepts. The developed *Disaster Management Data Model* has been based on the *Command and Control Information Exchange Data Model*. Although the *C2IEDM* is a sophisticated standard, the *DM²* points out important considerations that have to be taken into account for disaster management ontologies. The dynamic of the situation has to be treated with regard to temporal logic. Another important concept is context knowledge, which helps to fill semantic gaps. Finally, the nature of natural language text, has to be considered. Textual information contains always vagueness, which can be expressed in terms of uncertainty and reliability. Further research will be necessary to improve and further investigate the presented concepts in order to assemble a working prototype.

ACKNOWLEDGEMENTS

The presented work has been funded by the Deutsche Forschungsgemeinschaft (DFG), project no. BA 686/16 “Abstraction of Graphically and Verbally Represented Geoinformation” (Christian Lucas) and as part of the Collaborative Research Center (CRC) 461 “Strong Earthquakes: a Challenge for Geosciences and Civil Engineering” (Stefan Werder).

The authors would like to express their gratitude to the FGAN/FKIE for the provision of the SOKRATES ontology and the ongoing cooperation.

REFERENCES

Allen JF, Ferguson G (1994) Actions and Events in Interval Temporal Logic. Journal of Logic and Computation. Special Issue on Actions and Processes

Appelt DE, Israel DJ (1999) Introduction to Information Extraction Technology. Tutorial prepared for IJCAI-99. http://www.ai.sri.com/~appelt/ie-tutorial/IJCAI99.pdf

C2IEDM (2005) The C2 Information Exchange Data Model (C2IEDM Main). v. 6.15e. http://www.mip-site.org/publicsite/03-Baseline_2.0/C2IEDM-C2_Information_Exchange_Data_Model/

Cowie J, Lehnert W (1996) Information Extraction. Communications of the ACM 39(1):80-91

DV 100 (1999) Führung und Leitung im Einsatz, Vorschlag einer Dienstvorschrift DV 100. http://www.katastrophenvorsorge.de/pub/publications/DV100-SKK.pdf

DV 810 (1977) KatS-Dv 810 Sprechfunkdienst. http://gsb.download.bva.bund.de/BBK/KatS_Dv_810.pdf

Gerevini A, Schuber L (1994) An Efficient Method for Managing Disjunctions in Qualitative Temporal Reasoning. In: Proc. of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR94), Bonn, Germany

Gruber, TR (1993) A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5(2):199-220

Hecking M (2004) How to Represent the Content of Free-form Battlefield Reports. In: Proc. of the 2004 Command and Control Research and Technology Symposium, San Diego, California

JC3IEDM (2006) The Joint C3 Information Exchange Data Model (JC3IEDM Main). v. 3.1. http://www.mip-site.org/publicsite/04-Baseline_3.0/JC3IEDM-Joint_C3_Information_Exchange_Data_Model/

Kipper B, Jameson A (1994) Semantics and Pragmatics of Vague Probability Expressions. In: Proc. of the Sixteenth Annual Conference of the Cognitive Science Society, Atlanta, Georgia

Neumann G, Backofen R, Baur J, Becker M, Braun C (1997) An Information Extraction Core System for Real World German Text Processing. In: Proc. of the 5th International Conference of Applied Natural Language, Washington, USA

Neumann G, Declerck T (2001) Domain-Adaptive Information Extraction. In: Proc. of the International Workshop on Innovative Language Technology and Chinese Information Processing (ILT & CIP '01), Shanghai China

OASIS-CAP (2005) Common Alerting Protocol, v. 1.1. http://www.oasis-open.org/committees/download.php/15135/emergency-CAPv1.1-Corrected_DOM.pdf

OASIS-EDXL (2005) Emergency Data Exchange Language (EDXL) Standard Format for Resource Messaging. Draft v. 3.1. http://xml.coverpages.org/EDXL-Resource200508-14310.pdf

Pollard C, Sag IA (1994) Head-Driven Phrase Structure Grammar (Studies in Contemporary Linguistics). University of Chicago Press

Puppe F (1991) Einführung in Expertensysteme. Springer Verlag Berlin, Heidelberg, New York

Renooij S, Witteman C (1999) Talking Probabilities: Communicating Probabilistic Information with Words and Numbers. International Journal of Approximate Reasoning (22):195-215

Russell B (1923) Vagueness. Australian Journal of Psychology and Philosophy (1):84-92

Schade U, Frey M (2004) Beyond information extraction: the role of ontology in military report processing. In: Bachberger, E. (Ed.): Proc. of KONVENS 2004, Vienna, Austria

Stanford Medical Informatics (2007) The Protégé Ontology Editor and Knowledge Acquisition System. http://protege.stanford.edu

Werder S, Mueller M, Mueller M, Kaempf C (2006) Integrating Message Information into Disaster Management Maps: Transferability of a System of the Military Domain. In: Proc. of the ISPRS Commission IV Symposium on Geospatial Databases for Sustainable Development, Goa, India