Beyond Information Extraction:
The Role of Ontology in Military Report Processing

Ulrich SCHADE
FKIE, FGAN
Neuenahrer Straße 20
Wachtberg, Germany, 53340
schade@fgan.de

Miłosław FREY
FKIE, FGAN
Neuenahrer Straße 20
Wachtberg, Germany, 53340
m.frey@fgan.de

Abstract

Information extraction tools like SMES transform natural language into formal representation, e.g. into feature structures. Doing so, these tools exploit and apply linguistic knowledge about the syntactic and morphological regularities of the language used. However, these tools apply semantic as well as pragmatic knowledge only partially at best.

Automatic processing of military reports has to result in a visualization of the reports content by map as well as in an actualization of the underlying database in order to allow for the actualization of the common operational picture. Normally, however, the information provided by the result of the information extraction is not explicit enough for visualization processes and database insertions. This originates from the reports themselves that are elliptical, ambiguous, and vague. In order to overcome this obstacle, the situational context and thus semantic and pragmatic aspects have to be taken into account.

In the paper at hand, we present a system that uses an ontological module to integrate semantic and pragmatic knowledge. The result of the completion contains all the specifications to allow for a visualization of the report's content on a map as well as for a database actualization.

Introduction

In order to exploit Information Age concepts and technologies, the US Department of Defense proposes a transformation process aiming at "Network Centric Warfare" (2001). The basis of NCW is a network connecting deployed forces robustly such that they are able to exchange information easily. As a result, a common operational picture is supposed to emerge. Fleeting opportunities are to be recognized in time to take advantage of them, and thus speed, precision, and effectiveness are increased.

The theory of NCW has been developed to a large extent (Alberts and Hayes, 2003). It even has been evaluated in experiments (ibid). However, its implementation is still in the beginning. The problem does not lie in connecting the forces, physically. The problem is information itself. It raids the headquarters and their IT-systems in low quality and huge quantities. The quantity of incoming information overloads the capacities of the staff, and its low quality forestalls the emergence of the common operational picture. This is especially true for coalition forces due to a lack of interoperability.

1 Report Processing

This paper presents a system called the "SOKRATES" system which is meant to contribute to a better interoperability. The system processes military reports given in written natural language, say by SMS. A prototype has been completed in order to demonstrate the system's capabilities.

In its present form, the prototype is able to process reports of moving actions, e.g., "*Fünf Bradyland Haubitzen marschieren von Nederveert nach Helmond über Someren*" (*Five Bradyland howitzers moving from Nederveert to Helmond via Someren*) or "*Fünf feindliche Kampfpanzer in Zufahrt*" (*Five hostile battle tanks approaching)*. It also deals with "in position"-reports like "*Haben 31UFT785235 erreicht*" (*Arrived at 31UFT785235*). The input reports had been constructed on the base of real ones recorded during German army exercises.

Report processing within the system starts by splitting the report into header and content. The header of a military report provides at least the sender as well as a timestamp. It also may contain a reference to sender's location. The content is parted into sentence-like units. These units are transformed into a formal representation by means of Information Extraction (Appelt and Israel, 1999; Jackson and Moulinier, 2002, chapter 3). Automatic summarization (Mani and Maybury, 1999) is not sufficient. In order to meet the military demands the final representation of the report has to allow for specific post-processing. On the one hand, the content has to be integrated into the map displaying the common operational picture. On the other hand, the content has to be inserted into the data base. Because the formal representation normally does not include all the data needed for purposes it has to be augmented. This is done by ontological means.

2 Information Extraction

The information given by the reports in written natural language has to be represented formally. This is the task of information extraction. It is based on the SMES system (Neumann, 2003). The result of SMES-based information extraction is a feature structure (Shieber, 1986). In the case of report processing, it is a structure including an element of which "type" is the attribute and "report" the value. The structure also includes a mandatory element with attribute "sender." The value of "sender" is a structure of type "unit" representing the sender of the report. Thus, the sender's structure incorporates elements describing the relevant qualities of the unit in question, e.g., "name: 4./PzGrenBtl332, Zug C" and "size: PLT" (PLT for platoon). The content of the report is represented as value of "reporting_data" which is another mandatory attribute of the report's feature structure.

Details about the information extraction are given in Hecking (2003) but one aspect shall be mentioned here as well: The attributes of the feature structure as well as their values derive from the C2IEDM (Command and Control Information Exchange Data Model; cf. MIP 2003) used in NATO's Multilateral Interoperability Programme. By this programme, its member nations aim at interoperability of their command and control systems. The C2IEDM has been developed to code the essentials of battle space information in a uniform way. Therefore, the use of C2IEDM attributes and values allows for easy interaction with the data base accessible to all the forces deployed.

3 Augmentation

It is necessary to augment the result of the information extraction process for post-processing. In order to actualize the data base the inserts in question must respect constraints given by the C2IEDM declaring some data fields as mandatory. If the respective data is not in the feature structure after information extraction it has to be checked-up or calculated. Similarly, in order to visualize the report's content on a map, coordinates explicitly telling where to put symbols have to be provided. For example, if a report states that a symbol representing a unit has to be placed at the town of Nederveert, this can be done only if the coordinates of Neederveert are known.

All augmentation is done by ontological processing. The ontology is written in Protégé-2000. Its object hierarchy derived from the C2IEDM. E.g., it represented that a M1A2 (Abrams) is a battle tank, a battle tank is a tank, a tank is a vehicle, and a vehicle is an object. Instances are also defined as usual. E.g., a specific M1A2, identifiable by its call sign, is an instance of the type "M1A2 Abrams main battle tank." The ontology's domain is defined by scenarios in which military reports occur. For more details of ontology components in general, see McGuinness (2003); for details of the ontology under use, see Schade (2003).

C2IEDM information also is used for the definition of the objects' properties and their possible values. Placing restrictions on values is another step toward the representation of knowledge. E.g., "mobility" (a property of land vehicles) allows the values "tracked", "wheeled", and "towed." It is restricted to "tracked" in the case of M1A2s. However, the appropriate definition of the class hierarchy together with the definition of the classes' properties and a suitable restriction of their values is not enough. In order to augment the feature structures adequately, ontological processes (rules) must be added. These processes look up values or calculate them, sometimes even more complex operations have to be carried out. In the following, these cases are illustrated by example.

As a first example, we take the report "*Stellung bezogen*" (*in position*) sent by 2./PzGrenBtl332-ZugB. Analyzing this report, information extraction will construct a feature structure for sender which is of type unit. By the way, this feature structure is also the value of the reported action's agent. Thus, its symbol has to be displayed on the map during visualization. Information extraction provides the name of the unit (2./PzGrenBtl332-Zug B), and thus all the information needed to determine the unit's symbol according the APP-6A can be checked up: size = PLT, category = COMBAT, arm category = INF(mechanized), mobility = TRACKED, hostility = FRIEND. Even the coordinates of the unit's location can be checked up because the position deployed is represented in the data base.

The determination of values, especially, the determination of coordinates, however, is not always that simple. Let us assume, the 2./PzGrenBtl332-ZugB reconnoiters enemy movements and as a consequence reports "*Fünf Bradyland Haubitzen hinter Vinstedt*" (*Five Bradyland howitzers behind Vinstedt*). In this case, coordinates have to be calculated in order to place the howitzers' symbol on the map. It is safe to assume that the howitzers are "behind" the village of Vinstedt with respect to the position of the sender (secondary deitic reference). Thus, an axis is calculated from sender's location through the village of Vinstedt. The howitzers' symbol is placed on this axis next to Vinstedt, on the side where the sender is not.

A similar case is given if the unit reports "*Drei feindliche T80 in Zufahrt*" (*Three hostile T80s approaching)*. Again, coordinates have to be calculated. In this case, however, an intrinsic reference had been verbalized. The sender has an intrinsic orientation with respect to the front line, and it can be assumed that the enemy approaches from there. Thus, the coordination calculation is based on the specific scenario (e.g., the location of the front line), on military knowledge (e.g., the expectation that an enemy will approach from the direction of the front line) and on pragmatics, or to be more precisely, on the proper conduct of reporting and the proper referring to place.

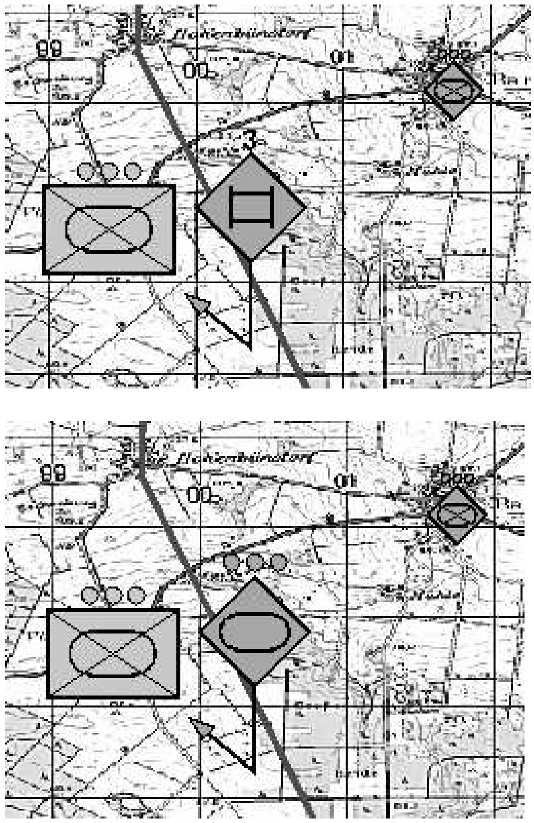

Figure 1: Visualization of the report "*Drei feindliche T80 in Zufahrt*" – upper part without and lower part with unit determination.

The last report ("*Drei feindliche T80 in Zufahrt*") may also serve as example for more complex processing which exploits mainly military knowledge. Whenever equipment is mentioned, a unit determination process may check which kinds of unit hold such an equipment. With respect to the example, it can be assumed that T80s will be operated by a battle tank unit, and because there are three of them the unit should be at least of platoon size. So, if the unit determination process runs, the visualization will display the symbol for "unit combat armor, hostile" (a red square with an oval inside) together with the size indicator for platoon (three dots) instead of the symbol for "equipment, armoured tank, hostile" (same red square with a kind of square inside) and the quantity indication ("3").

The unit determination process is facultative. The system's operator activates or inactivates facultative processes as required. In general, these processes automatize estimations which would be done more precisely by humans, but which sometimes have to be skipped due to a lack of time or resources.

4 Post-Processing

Post-processing means visualization of the report's content on a map as well as actualization of the underlying data base. With respect to visualization, all the units occurring in the augmented feature structure are displayed. In addition, context also is shown. The parts of the context to be shown (units, positions, barriers, control features) are determined by the operator and displayed according to the most actual knowledge about their locations. In order to highlight the report's content, its symbols are displayed somewhat larger than context's symbols.

The actualization of the underlying C2IEDM data base is done by a module which transforms the information kept in the report's feature structure into SQL-statements. Like the visualization module, it is coupled to the kernel of the SOKRATES system in a way that it can be deactivated or run on a different computer, e.g., together with the data base. The modular architecture of the SOKRATES systems grants that the system can be shaped as required. This flexibility assures adaptability to the actual situation and its demands. It also adds to military robustness.

Conclusion

IT systems to be used by coalition forces have to be interoperable such that these forces are enabled to exchange information, to cooperatively construct a common operational picture and to act in concert. Interoperability is the key ability for IT-systems used in coalition operations. The SOKRATES system can serve as a small piece of IT for this purpose.

The information extraction component of the SOKRATES system lays the foundation of its interoperability. Natural languages reports are transformed into a formal representation. However, this representation has to be augmented in order to allow for the intended post-processing, the actualization of the data base and the visualization of the report's content on the map. The augmentation is carried out by ontological processes. Only the whole process, information extraction and augmentation and post-processing, grants the level of interoperability demanded.

References

Alberts, D.S. and Hayes, R.E. (2003). *Power to the Edge*. CCRP, Washington, DC, 259 p.

Appelt, D. and Israel, D. (1999). *Introduction to Information Extraction Technology*. IJCAI-99, Stockholm, http://www.ai.sri.com/~appelt/ie-tutorial/ .

DoD (2001). *Network Centric Warfare. Department of Defense Report to Congress*, July 2001. Washington, DC.

Hecking, M. (2003). Information Extraction from Battlefield Reports. *Proceedings of the 8th IC2RTS*. CCRP, Washington, DC, 11 p.

Jackson, P. and Moulinier, I. (2002). *Natural Language Processing for Online Applications*. John Benjamins, Philadelphia, PA, 226 p.

McGuinness, D.L. (2003). Ontologies come of age. In: Fensel, D., Hendler, J., Lieberman, H. and Wahlster, W. (Eds.), *Spinning the Semantic Web*. The MIT Press, Cambridge, MA, 24 p.

Mani, I. and Maybury, M.T. (1999). *Advances in Automatic Text Summarization*. The MIT Press, Cambridge, MA, 434 p.

Multilateral Interoperability Programme (2003). *The C2 Information Exchange Data Model (C2IEDM)*, Working Paper 5-5, Version 6.1, http://www.mip-site.org.

Neumann, G. (2003).
http://www.dfki.de/~neumann/pd-smes/pdsmes.html .

Protégé-2000. http://protege.stanford.edu .

Schade, U. (2003). Towards an Ontology for Army Battle C2 Systems. *Proceedings of the 8th IC2RTS*. CCRP, Washington, DC, 10 p.

Shieber, S.M. (1986). An Introduction to Unification-Based Approaches to Grammar. CSLI, Stanford, CA, 105 p.