Computer Standards & Interfaces 36 (2013) 97–109

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Extending document models to incorporate semantic information for complex standards

Darren Clowes *, Ray Dawson, Steve Probets

Department of Computer Science, Loughborough University, UK

ARTICLE INFO

Article history:
Received 5 November 2012
Received in revised form 1 June 2013
Accepted 20 July 2013
Available online 31 July 2013

Keywords:
Semantic modelling
Information modelling
Aerospace
Defence
Quality

ABSTRACT

This paper presents the concept of hybrid semantic-document models to aid information management when using standards for complex technical domains. These standards are traditionally text based documents for human interpretation, but prose sections can often be ambiguous and can lead to discrepancies. Many organisations will produce semantic representations of the material. In developing these semantic representations, no relationship is maintained to the original prose. Maintaining the relationships has key benefits, including assessing conformance at a semantic level rather than prose, and enabling original content authors to explicitly define their intentions. This paper proposes a framework to help achieve these benefits.

1. Introduction

Increasingly the design of complex engineered products and systems (here within referred to as systems) is becoming more reliant on computer-supported models/representations of information, which can be used with computer checking algorithms to ensure consistency and correctness. Computer-supported models enable engineers to represent information graphically, which can aid understanding. These models also give a more rigorous definition of the system requirements than that of the more ambiguous (open to interpretation) prose representation of the same information. Many system properties can be modelled, such as, behaviour, functionality, verification information, and manufacturing instructions.

For the purpose of this research, a complex engineering domain refers to a domain where the documented materials are large in volume and contain elements that could be modelled semantically. These elements leave no possibility for different interpretations; as such they can be modelled to create the semantic information. Semantic information is the formalised description of the meaning and relationship between elements. This information is often represented in a semantic model. In particular, we use the term semantic model to refer to our models of the case study domain such as the Message Catalogue and Data Dictionary. This differs from document model which is a term used to describe a model which represents the structure of the document, i.e. paragraphs and tables. This model does not imply any knowledge of the information, more the spatial rendering of the information.

In engineering domains, the documentation is normally highly structured. Engineers are required to use, and have a detailed knowledge of these documents to perform their task. In complex engineering domains, these documents normally consist of large sections of prose. It can, therefore, be hard to understand the precise objective of the section. These sections generally contain elements which could be used semantically mixed with prose which aims to help the readers understanding, but can often be ambiguous.

There are many sectors of the complex engineering domain that could benefit from the use of semantic and static representation of information. The types of domains that will benefit include those that have associated standards documents. The defence industry contains many such documents. Other industries that have similar documents, which this technique may benefit, include the automotive, aerospace, nuclear and legislative sectors.

Most military standards in use are still disseminated as text-based documents. In our experience, this is also the case in most other domains requiring complex detailed standards. These standards can be thousands of pages in volume, which can make locating and composing information in them challenging and laborious. Within the development of military hardware, engineers are typically required to produce text based documents based on a subset of the standards implemented by their product. Product testing will require the engineers to promptly locate information in the standard and validate that their product conforms to it. Interoperability is also crucial in military applications where engineers must verify that their implementation does not impede or conflict with other products. To address these issues, we have found it useful to extract semantic models represented in a structured format which are then amenable to automated querying and processing. In our experience this approach can greatly enhance both the accuracy and speed of locating

* Corresponding author.
E-mail address: d.clowes@lboro.ac.uk (D. Clowes).

0920-5489/$ – see front matter
http://dx.doi.org/10.1016/j.csi.2013.07.009

and composing information from different parts of a standard. We have also found that in order to make models useful for engineers, it is essential to construct a suitable and familiar user interface for querying and navigating them.

2. Background

To evaluate the proposal of hybrid semantic-document models, the military domain of Tactical Data Links (TDL) was chosen as a suitable case study with the support of BAE Systems. Tactical Data Links display many of the characteristics of a complex engineering domain. The domain is specified through standard documents, each of which is many thousands of pages in length. They are highly structured and contain material that can be represented semantically. Regulation of conformance is obtained through production of documents similar in format and style to the original standard.

The TDL domain is a collection of related technologies designed to work within the command, control, communications, computers, and intelligence (C4i) used in the dissemination of information within a battlespace to support joint and combined operations. The TDL domain is often referred to as a family of standards, as several variations of Tactical Data Links have evolved to interface with specific unit types (infantry, aircraft, ships, etc.). These variations may differ in waveforms, bandwidths, protocols and capabilities.

The TDL provides one of the backbone technologies underpinning the defence community's goal of Network Enabled Capability (NEC) by providing the information and infrastructure to afford users with both an integrated picture of the battlefield and also provide tasking orders and responses. A number of TDLs are in service with coalition forces, and are implemented on a variety of assets, such as aircraft, ships, land vehicles and command stations, an example of which is illustrated in Fig. 1.

For the purposes of this project it has been decided to use the Link 16 standard. This is described in the U.S. Department of Defense Military Standard MIL-STD-6016C [1] and NATO Standardization Agreements (STANAGs) 5516 [2] and 4175 [3]. As both STANAGs are required for the complete representation of Link 16, the MIL-STD version will be the basis for this project's interpretation.

The Link 16 TDL is a general purpose TDL, in contrast to some others, e.g. Link 4A or Variable Message Format (VMF). A list of data link characteristics is provided elsewhere [4]. Link 16 has evolved over a number of years, stemming from a requirement identified by the US military in the early 1970s for a TDL offering a broad range of functions that would be applicable for use across multiple forces (e.g. Navy, Marines, Air Force, Army).

The Link 16 TDL (in MIL-STD and STANAG) is described in the form of narrative combined with many tables and relatively few figures. It is known to feature a number of shortcomings affecting its usability [5]. Of particular relevance to the research undertaken by the System Engineering Innovation Centre at Loughborough are, that the standard:

- is document-based, with no apparent underpinning model
- is largely narrative
- is open to (mis)interpretation
- is not checkable by machine
- contains duplication of material, inviting inconsistency
- has only very limited use of hyperlinking, impeding document navigability
- comprises many interdependent sections and appendices
- is enormous, >8000 pages.

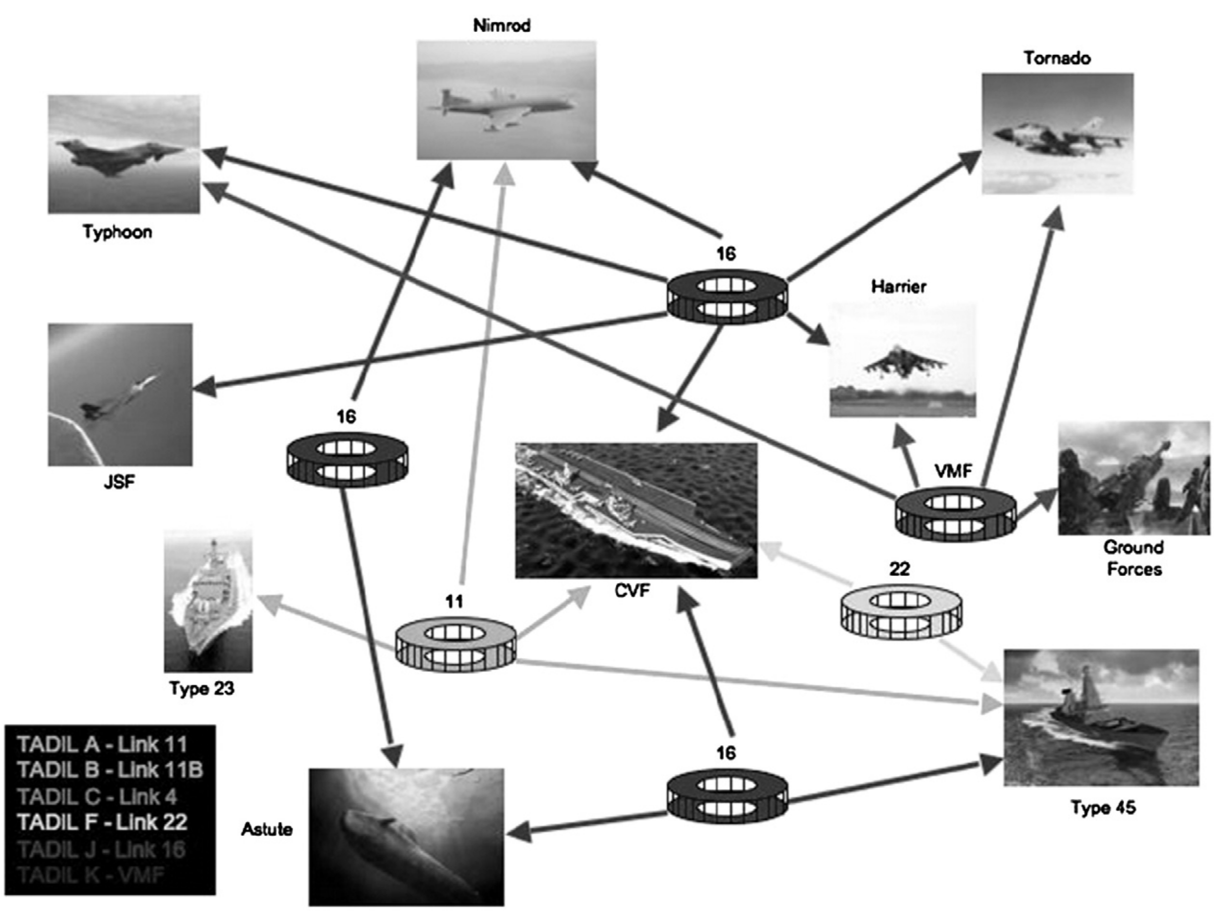

Fig. 1. TDL overview.