Knowledge Base L-V-C Mapping Method

Dong-Jae Lee, Yun-Hee Son, Kyu-Chul Lee
Chungnam National University
Department of Computer Science & Engineering
Daejeon, South Korea
e-mail: ehdwo115@gmail.com, {mellow211, kclee}@cnu.ac.kr

Abstract— **The defense training system uses an L (Live) system for practical training, V (Virtual) system for virtual training, and C (Constructive) system for combat command training. Recently, research to integrate the L-V-C training system has been under way to realize the same environment as an actual field. However, since the L-V-C integrated training system uses different middleware depending on the characteristics of L, V, and C, there is a problem with interoperability between middleware. The middleware used in each system is High-Level Architecture (HLA), Data Distribution Service (DDS), and Distribution Interactive Simulation (DIS). Each middleware uses a different data format: Federation Object Model (FOM), Topic, and Protocol Data Unit (PDU). In the case of FOM and PDU, there is a standard data format, but Topic does not specify a data format standard, so there is a problem with interoperability between heterogeneous middleware. In this paper, to solve the data interlocking problem of heterogeneous middleware, we constructed a knowledge base by extracting keywords based on the HLA FOM data format and extending it by ontology modeling. We also developed a knowledge base processing engine that supports interoperability between FOM and Topic using the built knowledge base and a weighted search algorithm.**

Keywords-LVC; Ontology; Knowledge Base; Keyword expansion; Keyword Extraction.

I. INTRODUCTION

The L-V-C integrated training system is a training system that integrates L (Live), V (Virtual), and C (Constructive) systems to perform various combat simulation exercises using virtual environments as real battlefields [1]. The L-system is real-life simulated training, which means it actually involves soldiers, military equipment, etc., operating in the real world rather than a virtual space. In the L system, the Data Distribution Service (DDS) is used as middleware because the real-time property is important. DDS is an OMG (Object Management Group) standard publishing / subscribing network communication middleware, and has the characteristics to support real-time [2]. V system means virtual training, which means that the soldier does not train in a real environment, but is training with real equipment in a virtual environment. The V system uses High-Level Architecture (HLA) as simulation middleware. HLA is a general-purpose architecture for distributed computer simulation systems [3]. Finally, the C system means combat command training, which means training with virtual

equipment and virtual forces in a virtual environment. The C system uses Distributed Interactive Simulation (DIS), which is a middleware running in a war game. DIS is an IEEE standard for performing war games in distributed locations while transmitting and receiving data messages in real time on a distributed network [4].

Recently, research on the L-V-C integrated training system has been conducted so that it can realize the same environment as an actual field. However, when integrating the L-V-C system, since each of the systems use different middleware, there is a problem with interoperability between middleware. The HLA, DIS middleware use FOM and PDU have different data format standards [4][5]. So, there is no difficulty mapping to each other. However, since Topic does not specify a data format standard, Topic is arbitrarily defined/published by the user. Therefore, problems arise when Topic is mapping with FOM and PDU, whose data format standards are defined.

One way to solve this problem is to map the Topic based on the standard FOM. However, users who do not know how the FOM is configured will have difficulty in mapping.

We solved this problem using ontology modeling and a knowledge base. By using the ontology to analyze the semantic keywords of the FOM and building the knowledge base through synonym based extension, users can map FOM and Topic through keyword search even if they do not know the exact FOM. In addition, there is an advantage, since overhead does not increase even if the complexity increases by using the ontology.

In this paper, we developed a knowledge base processing engine for interoperability HLA and DDS among the heterogeneous middleware HLA, DDS, and DIS used in the L-V-C. The contents of this paper are as follows.

- We analyzed and extended the semantic keywords of the HLA FOM with the standard for interoperability between HLA FOM and DDS Topic. In addition, we constructed knowledge base using ontology modeling.
- We developed a weighted search algorithm that allows users to search the associated keywords in priority order using weights in the knowledge base even if they are not familiar with HLA FOM.

The composition of this paper is as follows. In Section 2, we compare the related works. In Section 3, we explain the

background knowledge. In Section 4, we describe the knowledge base processing engine developed in this paper. Finally, Section 5 concludes the paper.

II.   RELATED RESEARCH

The works [6][7] are studies related to this paper. [6] is a study for combining HLA and DDS into a single middleware and [7] is a study for building a system that can utilize both HLA and DDS. Both studies map only the data defined in the HLA standard and the data in the DDS specification to link HLA and DDS. This creates mapping difficulties when users add a new Topic. In this study, we support mapping between HLA FOM data and Topic, even if the user adds a new Topic.

III.   BACKGROUND

A.   Ontology model

In this paper, we used Resource Description Framework(RDF) of N-Triple format to construct the ontology model [8]. RDF is a World Wide Web Consortium(W3C) standard technology that provides interoperability between applications that exchange information that is machine understandable on the Web [9]. In addition, Simple Protocol and RDF Query Language (SPARQL) was used to query the constructed knowledge base. SPARQL is a database query language that can search and manipulate data stored in RDF. In addition, SPARQL is recognized as one of the key technologies in the Semantic Web with the standard technology established by the W3C [10]. In this paper, we employed Jena [11] to use RDF and SPARQL. Jena is a Java-based open source semantic web framework and provides a programming environment for RDF, RDF Schema(RDFS), Web Ontology Language (OWL), SPARQL, and Rule-based reasoning engines. Finally, in this paper, Jena TDB was used to store the constructed model in the knowledge base, and Jena TDB provided the function to store and manage the RDF format data [12].

B.   Keyword Extraction, Extension

In this paper, Natural Language Processing(NLP) was used to extract nouns from the semantics of HLA FOM. OpenNLP, one of NLP's open source is a machine learning based tool for natural language text processing. It supports most common NLP tasks such as tokenization, sentence segmentation, part of speech tagging, entity extraction, and parsing, and supports advanced text processing services [13]. We also used WordNet to extend the extracted keywords. WordNet is a database in which about 150,000 words including nouns, verbs, adjectives, and adverbs are stored in a set of 115,000 synonyms [14].

C.   HLA FOM and DDS Topic

HLA FOM is a set of federated objectives, which is an IEEE standard that contains a specification that describes the shared objects class and objects class' name, object class attributes, and interactions of the federation [5].

A data model is a description of the state of a system, including data types, processes for data transfer, and data access methods. DDS operates as defined by this data model, using the Global Data Space. DDS Topic is used to identify data in Global Data Space, and the user defines the Topic directly according to the data model [15].

D.   Interoperability

Research on interoperability has been increasing since 1970. Interoperability is used in a variety of areas and there are 34 different definitions mentioned in research papers, standards and government documents over the past 30 years [16]. Among the various definitions, the definition that corresponds to the system we developed is "The ability of two or more systems or components to exchange and use the exchanged information in a heterogeneous network "[17].

IV.   KNOWLEDGE BASE PROCESSING ENGINE

A.   Knowledge base composition diagram

Figure 1 shows the overall structure of the LVC knowledge base processing engine. In order to map the FOM and Topic that the user arbitrarily defines, the knowledge base processing engine processes the total of two processes. One is the process of building a knowledge base. It extracts keywords from FOM through a keyword extraction process and expands extracted keywords through a keyword expansion process. The extended keyword is converted into the N-Triple format through the ontology modeling process and stored in the knowledge base (triple store).
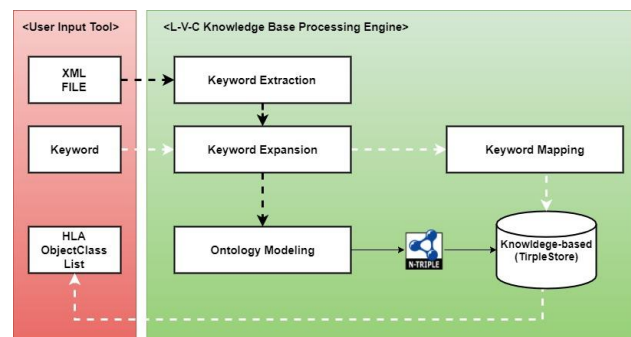

Figure 1. Overall structure of the LVC knowledge base processing engine

The other process is the keyword search process. We expanded the FOM keyword in the knowledge base building stage. However, to further increase accuracy, we also expanded the keywords that the user types and then performed the keyword search process. In the keyword search process, after querying the knowledge base using the weighted search algorithm, the extended keyword is returned and a list of the object class to which the mapping is performed is returned.

B. Knowledge base construction

Although existing HLA and DDS mapping methods can be mapped only to predefined ones, it is necessary to build a knowledge base to map the Topic and HLA that the user defined arbitrarily. We constructed the knowledge base using synonyms to support user convenience. The knowledge base construction is divided into three stages: keyword extraction, keyword expansion, and ontology modeling.

1) Keyword Extraction: In the keyword extraction process, the HLA FOM (XML) file is received through user input. It parses the semantic sentence describing the object class name and object class in the FOM and uses OpenNLP to extract nouns from object class names and semantic sentences.

Figure 2. Example of FOM file

Figure 2 shows an example of an HLA FOM file. The name of the object class is 'Aircraft' and there is a semantic sentence describing it. The object class of HLA FOM is described in the background section.

Figure 3. Aircraft extraction result

Figure 3 shows the nouns extracted from the object class names and semantic sentences in the FOM File in Figure 2 using OpenNLP. The object class name 'Aircraft' and the nouns 'platform', 'entity', 'air', 'aircraft', 'entities', and 'ground' in the semantic sentence are extracted.

2) Keyword expansion: To extract keywords based on synonyms, the extracted nouns are searched in a separate dictionary built in WordNet. When a retrieved noun is searched in the Wordnet dictionary, the synonym result for the extracted noun comes out. This result extends the data. Figure 4 shows the expanded word 'platform' extracted from Figure 3. 'Platform' expands to 'platform', 'political_platform', 'political_program', 'program', 'chopine' .

3) Ontology Modeling: Figure 5 shows the ontology model. The object class has subclasses as Keyword, and the Keywords are subclasses that have an Expansion Keyword extended through WordNet. Also, the Keyword and the Expansion Keyword have a weight indicating their priority.
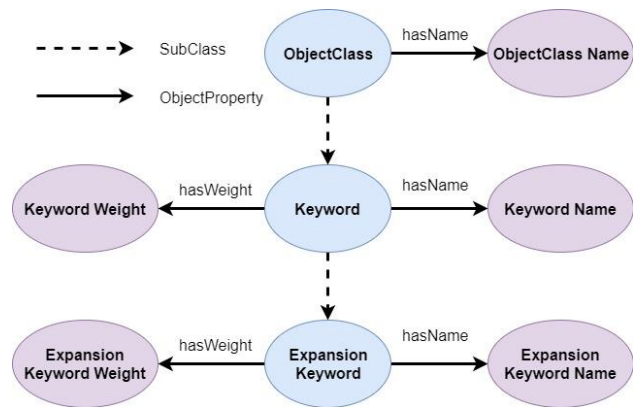

Figure 4. Example of Platform extension

Figure 5. Knowledge base ontology model

Data generated by the ontology modeling is stored in N-Triple format and the created N-Triple is stored in the database using Jena TDB.

C. Keyword search

The Keyword search performs the SPARQL query by receiving the keywords (e.g., 'craft') necessary for the ontology query, based on the knowledge base constructed above. When performing a query, one you may not know exactly what object class name to look for. Because the knowledge base is built by synonym extension, in this case one can still retrieve the associated object class name. In addition, when the keyword alone does not produce a result, the input keyword is further expanded by synonyms to query the constructed knowledge base to derive the result. For this process, weighted search algorithms were newly implemented and used.

D. Weighted search algorithm.

In Figure 5, Keyword and Expansion Keyword have weighted properties. The method of weighting is as follows

- Keywords are extracted from the object class name are assigned a weight of 1, and keywords are extracted from a semantic sentence are assigned a weight of 2.

- In the above process, expanded keywords are re-expanded through WordNet, and the weight is incremented by one in the expanded keywords order.

The object class is organized into a tree using these weights. Figure 6 shows an example of a tree.
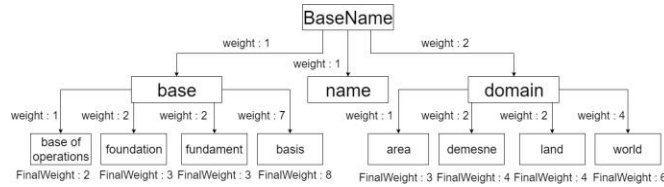

Figure 6. BaseName tree example

This tree is as large as the number of ObjectClass, and is sequentially searched in three cases using the configured tree as follows.

1) Keyword (knowledge base) and Keyword (user search) mapping: The keyword of the knowledge base is mapped to the keyword inputted by the user.

2) Expansion Keyword (knowledge base) and Keyword (user search) mapping: The expansion keyword of the knowledge base is mapped to the keyword inputted by the user.

3) Expansion keyword (knowledge base) and Expansion keyword (user search) mapping: The expansion keyword of the knowledge base is mapped to the extended keyword inputted by the user.

At each step when a result is obtained, it ends without going to the next step. If the keyword entered by the user in each search step matches the keywords stored in the plurality of *object class trees, the weighted* values are compared and the keywords of the lowest weighted value are output in ascending order from the keywords of the tree.

V.　CONCLUSION AND FUTURE WORK

In this paper, we analyzed related technologies needed to develop a knowledge base processing engine. We analyzed OpenNLP and WordNet for keyword extraction and extension of the L-V-C knowledge base processing engine. We also analyzed the ontology representation language, RDF, and the query language, SPARQL, and analyzed the Jena TDB for ontology data storage. Based on these related technologies, we developed a knowledge base processing engine that supports interoperability between heterogeneous middleware.

Through this study, it was confirmed that more flexible interoperability and expansion is possible knowledge base of

ontology based . Also, the development of a knowledge base processing engine enables data interoperability through keyword-based retrieval even if there is no prior knowledge of the L-V-C system. Therefore, it is expected that users' barriers to entry will also be lowered, and the knowledge base can be utilized in various fields.

It is expected that future research plan will be able to develop a knowledge base processing engine that is not dependent on a specific field so that it can interoperate with the heterogeneous middleware of L-V-C system as well as through knowledge base in various fields.

REFERENCES

[1]　B. Pollock, E. Winer and S. Gilbert, "LVC interaction within a mixed-reality training system." The Engineering Reality of Virtual Reality 2012. Vol. 8289. International Society for Optics and Photonics, 2012.

[2]　G. Pardo-Castellote, "Omg data-distribution service: Architectural overview," Proceedings of 23rd IEEE International Conference on Distributed Computing Systems Workshops, pp. 200-206, 2003.

[3]　J. S.dahmann, "High Level Architecture for Simulation," Proceedings of the First International Workshop on Distributed Interactive Simulation and Real-Time Application, pp. 9-14, 1997.

[4]　DIS Steering Committee, "IEEE standard for distributed interactive simulation-application protocols." IEEE Standard 1278, pp 1-52, 1998

[5]　HLA Working Group, "IEEE standard for modeling and simulation (M&S) high level architecture (HLA)-framework and rules." IEEE Standard,1516-2000, 2000.

[6]　P. Yunjung and M. Dugki, "Development of HLA-DDS wrapper API for network-controllable distributed simulation," Application of Information and Communication Technologies (AICT), 2013 7th International Conference on. IEEE, pp. 1-5, 2013.

[7]　J. Rajive and G. Pardo-Castellote, "A comparison and mapping of data distribution service and high-level architecture," Technology, The Netherlands. His research interests include parallel and distributed computing, component based architectures, and embedded systems, 2006.

[8]　O. Lassila and R. R. Swick, " Resource description framework (RDF) model and syntax specification,", 1999.

[9]　World Wide Web Consortium. "About the World Wide Web Consortium (W3C).", 2001.

[10] E. Prud and A. Seaborne, "SPARQL query language for RDF," 2006.

[11] Jena, Apache, "semantic web framework for Java," 2007.

[12] A. Seaborne, "Jena TDB," 2011

[13] J. Baldridge and T. Morton, "OpenNLP," 2004

[14] C. Fellbaum, WordNet John Wiley & Sons, Inc., 1998

[15] G. Pardo-Castellote, "Data-Centric Programming Best Practices: Using DDS to Integrate Real-World Systems.", 2010.

[16] C. Ford-Thomas, et al., "Survey on Interoperability Measurement", AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH, 2007.

[17] A. Geraci et al., "IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries", IEEE Press, 1991.