

한국형 임무공간 개념모델 일관성 검증기법 적용 사례

김정윤^{*,1)} · 이경화²⁾

¹⁾ 휴먼C&C M&S 연구실

²⁾ 쌍용정보통신 국방사업부

A Case of Consistency Verification Technique on Korean CMMS

Jungyoon Kim^{*,1)} · Kyunghwa Lee²⁾

¹⁾ M&S Lab., Human C&C, Korea

²⁾ National Defense Dept., SSangYong Information & Communication, Korea

(Received 19 August 2024 / Revised 28 October 2024 / Accepted 6 November 2024)

Abstract

The Concept Model of Mission Space(CMMS) can be regarded as an ontology that systematically represents knowledge within a military domain. Ensuring the consistency of this ontology is crucial, and its consistency must be verifiable. This paper presents a case study on the consistency verification of the Korean CMMS, referred to as CMMS-K. The verification feature can detect inconsistencies such as duplication, missing links, and circular definitions within the ontology elements. This capability is achieved through the formality of the Methontology template, which provides a structured specification of ontology elements. The implemented feature demonstrates the practical ability to perform verification and highlights the future prospects of the Korean CMMS.

Key Words : CMMS(임무공간개념모델), Ontology(온톨로지), Consistency(일관성), Verification(검증), M&S

1. 서론

군(軍) 온톨로지는 군(軍)의 정보공유와 지식표현을 쉽게 함으로써 신속하고 효과적인 지휘결심에 기여한다. 규칙에 따라 체계적으로 표현된 온톨로지는 다양한 분야에 다양한 목적으로 활용될 수 있다. 이미 해외선진국의 군은 지휘통제(Command & Control(C2)), 정보감시정찰(Intelligence, Surveillance, and Reconnaissance),

모의훈련(Simulation and Training), 군수(Logistics), 정보보호(Cybersecurity) 등에 온톨로지를 적용하고 있다^[1]. 특히 모의모델을 개발하기 위한 표준지식을 제공함으로써 모의모델 중복개발(stovepipe)과 같은 문제를 경감시킨다. 온톨로지는 단순한 지식사전의 수준을 넘어서 지식 그 자체의 일관성(Consistency)을 확인할 수 있는 규칙(syntax)을 제공함으로써 지식의 비일관성(Inconsistency)을 자동적으로 식별할 수 있는데, 이는 모의모델의 검증(verification)과 확인(validation)을 위한 목적으로 활용될 수 있다.

한국형 임무공간개념모델(Korean Concept Model of

* Corresponding author, E-mail: jkim9@kaist.ac.kr

Copyright © The Korea Institute of Military Science and Technology

Mission Space, CMMS-K)은 군 온톨로지로서, 육해공군 및 해병대의 전술적 교리가 반영된 임무공간을 개념모델로 표현하여 데이터로 저장하고 활용하는 것을 목표로 한다. 이 개념모델은 모의논리의 개발을 위한 기반지식이다. CMMS-K는 개념모델의 일관성을 검증할 수 있는 기능을 포함한다.

CMMS-K는 Fig. 1에 표현된 것과 같이 일관성(consistency) 검증과 정확성(correctness) 검증의 두 기능을 포함하는데, 각각 프로그램의 문법(syntax)과 의미(semantics) 검사와 비교될 수 있다. 즉, 컴파일러는 프로그램 코드의 문법 오류(syntactic error, compile time error)를 식별하지만 의미오류(semantic error, runtime error)는 식별하지 못한다. 의미오류는 디버깅을 통해 동적으로 파악할 수 있을 뿐이다. 마찬가지로 온톨로지도 오류를 자동적으로 식별할 수 있는 부분은 규칙 오류를 식별하는 일관성 검증이며, 의미의 오류를 식별하는 정확성 검증은 도메인전문가에 의해서만 가능한 것으로 본 논문에서는 구분한다. 본 논문은 첫 번째의 일관성 검증을 어떻게 구체화했는지 그 적용 사례를 소개한다.

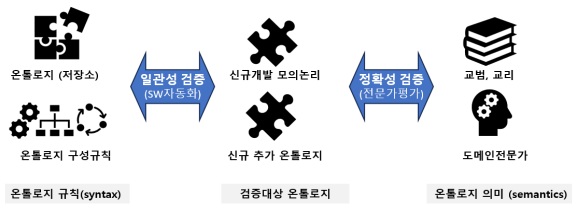


Fig. 1. Ontology consistency and correctness

CMMS-K의 규칙은 Methontology^[2]로부터 획득되었다. Methontology는 온톨로지를 작성하기 위한 방법론으로 총 11단계의 절차(process)와 양식(template)을 제공한다. 각 단계에서 개념요소가 정의되고 해당 요소와 관련된 속성과 이진관계, 공리, 규칙, 상수, 인스턴스가 양식에 작성된다. 절차와 양식은 온톨로지 구성요소가 상호 연계될 수 있도록 규칙에 맞추어 작성되며 결과적으로 온톨로지 구성의 정형성(formality)을 확보한다. 일관성 검증은 이 정형성을 활용한다.

CMMS-K의 일관성 검증은 M&S 체계 개발과 모의논리 검증 및 확인(Verification and validation, V&V)을 위한 가이드라인으로 활용될 수 있다. 또한 온톨로지 자체 품질확보를 위해서도 이 기능이 필요하다. 본 연구는 온톨로지 구성요소의 단계 간 불일치(missinmg

links), 용어중복(redundancy), 순환정의(circular definition) 오류, 표기오류(typo)를 식별하기 위한 도구 개발을 포함하며, 본 논문은 도구개발 사례를 소개한다.

한편, 정확성(correctness) 검증은 도메인전문가 평가를 지원하는 기능으로 구현하였는데, 그 내용은 본 논문의 범위를 벗어난다.

2. 온톨로지와 검증기술

2.1 온톨로지

온톨로지(Ontology)는 ‘존재론’이라 하며 원래 사물의 존재 의미를 논의하는 철학적인 연구 영역을 뜻하는 말이다. 본 논문의 온톨로지(Ontology)는 이해당사자(stakeholders)들 간에 합의를 이룬 바를 개념적이고 컴퓨터에서 다룰 수 있는 형태로 표현한 모델로, 개념의 타입이나 사용상의 제약조건들을 명시적으로 정의한 지식표현(knowledge representation)이다. 따라서 컴퓨터가 온톨로지 표현의 개념을 이해하고 지식처리를 할 수 있게 된다. 온톨로지는 프로그램과 인간이 대상을 도메인의 지식을 공유하며 도메인 내 개념을 명확하게 정의하고 기술하여 정확한 정보를 찾을 수 있도록 한다.

2.2 온톨로지 구성요소 간 관계

온톨로지는 그 구성요소가 상호 연관성을 가지는데, 이 연관성은 온톨로지의 일관성 검증(consistency checking)을 위한 중요한 수단이 된다. 일관성 검증은 온톨로지에 오류, 모순 또는 부정확함으로 인해 문제가 없는지 확인하고 자동화된 추론(automated reasoners)의 기본 바탕이 된다^[3]. Fig. 2는 개체관계에서의 추론예시를 보여준다. 화살표는 일반화(is-a)를 말하는데, Parent와 Woman은 Person으로 일반화 되고, Woman은 Person과 Female의 공통속성을 상속받는다. Female로서 Parent의 속성을 가진다면 당연히 Mother가 될 것이다. 오른쪽의 hasChild 씨클은 Parent 역할(Role)을 갖기 위해서는 Person이 가져야 할 제한사항으로서 value restriction으로서 (1, NIL)을 부여하고 있다. 이는 Person이 하나(1) 이상(NIL = infinity)의 아이가 있어야 (hasChild) Parent가 될 수 있다는 뜻이다. 실질적인 검증을 위한 기술 및 도구들이 제시되고 있는데, 다음은 온톨로지 일관성을 수행하기 위한 몇 가지 일반적인 기술이다.

여기서 개체관계가 적절한 관계를 갖추는 것을 정확성 관련 문제로 표현하고 있는데^[3], 본 논문은 이러한 관계도 일관성 측면의 문제만 고려한다. 당연히 Person이 Woman의 일반화된 관계임을 우리는 상식으로 알지만, 어떤 도메인 내에서 개체 용어가 두 개가 있을 때 어느 쪽이 더 일반화된 개념인지 판단하는 것은 컴퓨터가 할 수 있는 내용이 아니다. 하나의 용어가 일반화된 개념으로 판단되었으면 다른 하나의 용어는 그 용어의 상위에 있을 수 없다(순환정의 오류). 일관성 검증은 이러한 측면의 문제만 판단한다. 즉 그 두 용어가 모두 상위관계에 있다고 표현되었을 경우 일관적이지 못한 상태(inconsistent)가 된다.

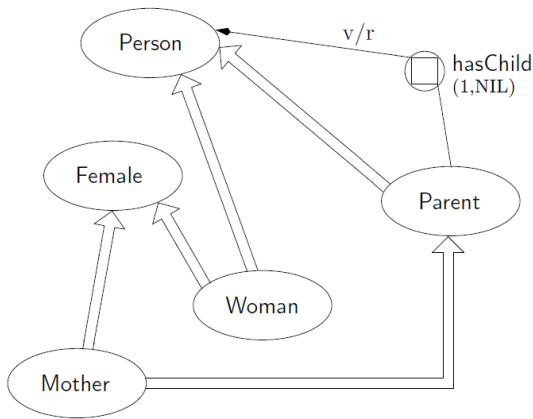


Fig. 2. Reasoning of entities and relations

2.3 온톨로지 일관성 검증

온톨로지 일관성 검증은 온톨로지 지식표현 내용의 요소간 비일관성을 평가한다^[3]. 온톨로지를 구성하는 핵심 지식표현은 클래스와 속성인데, 온톨로지의 일관성이 높다는 것은 일관성 오류가 발생한 클래스 및 속성의 비율이 낮다는 의미이다. 다음은 여러 연구내용에서 제시되는 일관성 오류 발생 유형이다.

용어 중복^[4] : 온톨로지 내 같은 용어를 가지는 서로 다른 클래스가 존재하는 경우, 그리고 같은 인스턴스를 가진 의미적으로 동일한 서로 다른 클래스가 존재하는 경우, 서로 동시에 존재할 수 없는 클래스가 같은 서브클래스를 가지는 경우(Fig. 3), 온톨로지 명명 규칙에 따라 정의되지 않은 용어가 존재하는 경우이다. Fig.3.을 보면, Class-B는 p1과 p2의 서브클래스로 정의되어 있다. 하지만, p1과 p2는 disjoint class이므로 서로 동시에 존재할 수 없는 클래스이다. 따라서

이는 클래스 정의에 오류가 있음을 알 수 있다. 예를 들면, 개(p1)와 고양이(p2)는 포유류(A)라는 클래스의 disjoint 클래스이지만, 도베르만(B)이라는 클래스가 동시에 p1과 p2의 서브클래스로 정의되어 있다면 일관성 오류이다.

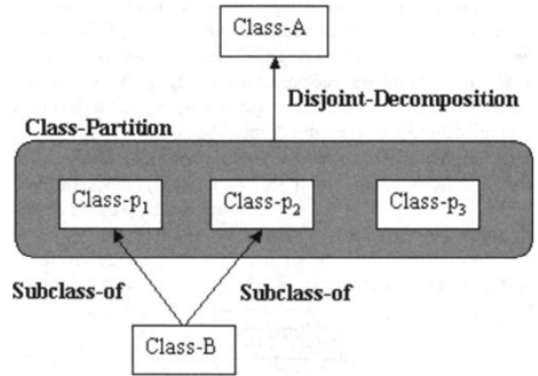


Fig. 3. Inconsistency of joint and disjoint

구조적 오류^[4] : 계층 구조에 순환 구조가 적용되어 있는 경우이다. Fig. 4는 구조적 오류의 사례를 보여주고 있다. 논리적으로 “A는 B이고, B는 C인데, C가 다시 A”라는 식으로 자기 모순을 갖고 있는 경우이다. 본 논문에서는 이 내용을 순환정의(circular definition)로 표현한다.

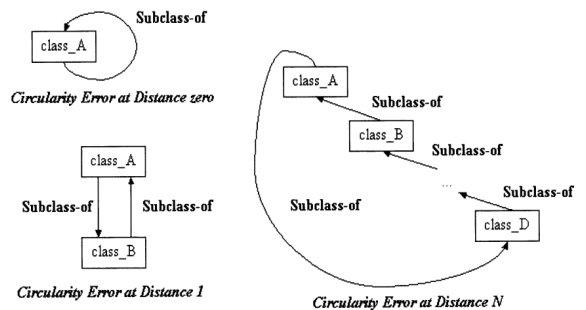


Fig. 4. Circular inconsistency

의미적 오류^[5] : 구조적 문제는 없으나, 의미적으로 계층구조나 관계가 적절치 못한 경우이다. 그런데, 적절치 못한 관계인 것을 판단하는 것은 사람의 개입 없이 이루어지기 어렵다. 의미적 문제의 예로서, 의미상 house는 month-name에 속할 수 없다. 집을 의미하는 house는 month-name과 관계가 없기 때문이다.

전체 일관성 오류^[5] : 구조적인 문제 및 의미적인 문제가 없으나, 전체 온톨로지 일관성을 해치는 경우이다. 의미적 문제의 경우와 마찬가지로 역시 사람의 개입 없이 판단하기 어려운 오류이다. 전체 일관성 문제는 의미가 같지만 표현이 다른 경우도 존재하는데, 이는 언어 간 차이 등에 기인한다. 스페인어로 1월을 의미하는 Enero는 의미적으로 month-name과 관계가 존재하지만, 전체 온톨로지 일관성 측면에서 영어가 아니므로 적절치 않은 경우로 예를 들 수 있다^[5].

본 연구에서는 용어중복과 순환정의의 일관성 검증을 다룬다. 의미적 오류와 전체 일관성 오류는 CMMS-K의 정확성 검증 내용에 해당하므로 일관성 검증 대상으로 하지 않는다.

3. 한국형 임무공간 개념모델(CMMS-K)

국내외 학계와 다양한 사회기관, 선진국 군(軍)에서 명칭은 다양하지만 온톨로지 또는 사실상 온톨로지에 해당되는 데이터 모델을 활용하고 있다.^[1] 한국군의 경우 온톨로지 역할을 하는 무기체계 세부분류를 포함한 다양한 분류자료가 있으나, 컴퓨터 환경에 통합 구축되어 온톨로지와 같이 정보화 지식체계로서 정보 제공을 한다는 측면에서 미흡한 면이 있다. 따라서 한국군도 자체 도메인에 대한 온톨로지 구축 필요성이 있으며, 이에 따라 한국형 온톨로지 개발을 추진하고 있다. 본 논문은 그러한 노력의 일환이다.

CMMS-K는 향후 한국군의 모의모델 개발을 위한 지식 획득과, 또한 다양한 한국군 모의모델의 표준 제공에 중점을 둔다. CMMS-K 연구는 도메인 텍스트자료(교범, 교리, 기존의 모의논리서 등)로부터 임무공간 개념모델을 자동적으로 생성하는 기술, 생성되거나 작성된 온톨로지를 검증하는 기술을 포함한다. 이를 위해 본 연구는 한국형 온톨로지 방법론을 정립하여 이를 바탕으로 온톨로지가 정형성(formality)을 갖도록 한다.

한국형 온톨로지 방법론은 온톨로지 요소를 명세하기 위한 양식을 정의하며, 그 작성 방법에 대한 안내이다. 한국형 온톨로지 작성 양식은 Methontology^[2]를 기반으로 하며, 한국형 임무공간 개념모델을 구축하기 위해 Methontology 방법론을 맞춤화(tailoring)했다. 구체적인 양식과 작성 절차는 Ontology Engineering의 “3.3.5 METHONTOLOGY”^[6]를 참조했으며 그 절차는 Fig. 5와 같다. Methontology는 도메인개체(entity) 정의

와 지식획득(knowledge acquisition) 두 활동과, 온톨로지 구성요소(components) 설정에서부터 시작된다. 앞의 두 활동은 온톨로지 자체이고 구성요소는 온톨로지를 표현하는 방법이다. 특히 구성요소는 온톨로지의 일관성 검증(consistency checking) 및 정확성 검증(correctness checking)의 중요한 수단이 되는데 이를 바탕으로 자동추론기(automated reasoner), 정형방법(formal method), 전문가 검토(expert review), 그리고 온톨로지 디버깅 도구(ontology debugging tools)와 같은 기술 적용을 위한 기반이 된다^[7].

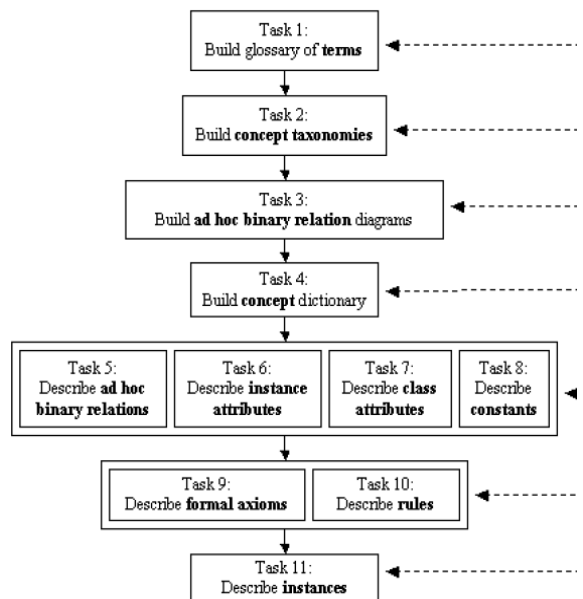


Fig. 5. Methontology phases

Fig. 6은 Methontology를 CMMS-K에 맞게 맞춤화한 방법론이다. 이 방법론에서 도메인 온톨로지는 개체(Entity) 측면과 과업(Task) 측면이 동시에 개발된다. 예를 들어, 일반적으로 도메인 내에서 개체는 과업을 수행하는 주체나 수단이 될 수도 있고 과업은 개체에 의해 진행되는 절차가 될 수도 있다. 따라서 본 논문은 구조적인 요소(structural elements)인 개체와 행위적 요소(behavioral elements)인 과업 측면에서 구분하였으나 두 요소가 밀접하게 연계되어 명세된다. 즉 개체를 다룰 때 개체가 연루된 과업도 염두에 두고 명세(specification)한다는 것이다. 그런데 지식체계(hierarchy)와 명세규칙(syntax) 상 일관성 측면에서 유사하기 때문에 개체 사례만 본 논문에서 설명한다.

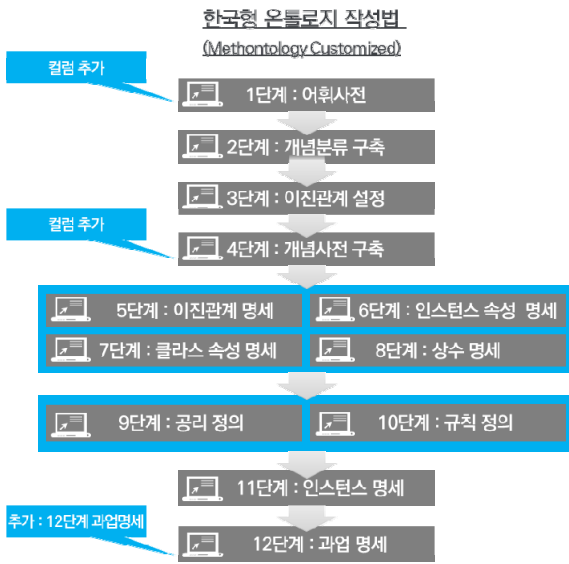


Fig. 6. Tailored methontology for CMMS-K

일반적인 온톨로지 사례에서 개체와 과업은 다음의 의도(purpose)로 작성된다. 개체의 경우 개체 연결(linking) 및 명확화(disambiguation), 온톨로지 모듈화 및 재사용(modularization and reuse), 평가 및 검증(evaluation and validation), 그리고 시각화 및 탐색(visualization and exploration)을 위해 작성된다⁷⁾. 과업의 경우 이벤트 모델링 및 표현(modeling & representation), 이벤트 기반 시스템(event driven system), 온톨로지 기반 이벤트 처리(ontology based event processing), 그리고 이벤트 기반 지식관리(event based knowledge management)를 위해 작성된다⁸⁾. 한국형 온톨로지 방법론에서의 개체와 과업도 이러한 의도를 따른다.

방법론의 산출물은 “온톨로지정의서”로 명명하였으며, 여러 양식(template)을 엑셀파일의 시트로 구체화되어 있다. 총 12단계의 온톨로지정의서는 1단계 어휘사전, 4단계 개념사전, 5단계 이진관계, 6단계 인스턴스 속성, 7단계 클래스 속성, 8단계 상수목록, 9단계 공리정의, 10단계 규칙정의, 11단계 인스턴스 목록, 12단계 과업목록이라는 산출물로 구성된다. 2-3단계는 도메인전문가가 파워포인트나 종이 위에 개념을 그려서 시각화하는 작업단계이므로 양식으로 정의되지 않는다. 이 단계에서 시각화한 내용은 5단계 이진관계에 정리된다. 각 단계의 양식은 각각 규칙이 있으며 이 규칙에 따라서 내용이 명세됨으로써 정형성(formality)을 갖추게 되고, 검증을 위한 기반을 제공한다.

4. CMMS-K에 적용된 검증방법

CMMS-K의 일관성 검증은 개념모델의 개념 중복성, 불일치, 순환정의, 표기오류를 자동적으로 식별하는 것을 목표로 한다. 내용은 다음과 같다.

- **중복성(duplication)** : 중복된 개념용어를 식별한다. 어휘사전의 어휘는 유일(Unique)한 하나의 집합(set)이다. 1단계 어휘는 어휘사전 내에서는 유일하다. 그런데 어휘와 어휘간의 조합이 유일한 상황도 있다. 어휘가 4단계 이후에 명시될 때는 개념, 이진관계, 속성, 등으로 분류되어 명시되는데 이 단계에서는 하나의 어휘가 다른 어휘와의 관계를 가지기 때문에 4단계 이후의 개념사전, 이진관계정의, 속성정의 등의 양식에서 작성될 때는 어휘가 중복될 수 있다. 이 양식에서는 중복되는 어휘는 다른 어휘와 조합됨으로써 그 조합의 유일성을 유지해야 한다. 예를 들어 ‘항공기’ 개념은 ‘속도’와 ‘탑재량’을 속성으로 가지는데, 두 상황을 개념사전에 표현하면 ‘항공기-속도’, ‘항공기-탑재량’이라는 조합으로 된다. 이때 ‘항공기’가 두 번 중복되어 표현되지만 조합이 유일속성이 된다.
- **불일치(missing link)** : 단계 간 일치되지 않은 용어를 식별한다. 어떤 어휘가 어휘사전에서 개념, 이진관계, 속성 등으로 구분되었다면, 당연히 해당 양식에 명시되어야 한다. 1단계에서 어휘사전에 등록된 어휘는 어휘의 유형(type)에 따라 뒤쪽 단계의 해당 양식에 나타나야 한다는 것이다. 예를 들어, 개념은 개념사전, 이진관계는 이진관계정의, 속성은 속성정의 양식에 명세되어야 한다. 아니면 어휘사전이나 해당 양식에 누락된 상태이다.
- **순환정의(circular relation)** : 상하 관계 중 상위개념이 하위개념의 하위개념(또는 그 반대)에 위치하는 순환적 정의 모순을 식별하는 것이다. 이진관계 중 상하 관계의 성격이 있는 관계에서 상위개념이 하위개념의 하위에 위치하거나 하위개념이 상위개념의 상위에 위치하는 경우이다. 상하위 관계의 이진관계(is-a, has-a, is-affiliated-with, has-a-part-of, 등)가 모두 해당될 수 있다.
- **표기 오류 typo)** : 철자 및 띄어쓰기 오류를 식별한다. 정형화되어야 하는 온톨로지 특성상 사소한 철자 오류나 띄어쓰기 차이는 다른 어휘로 간주한다.

유사한 용어를 식별해서 제시하거나 같은 의미의 용어가 띄어쓰기만 다른 경우도 제시한다.

- 서브모델(하위개념) 간 차이 식별(sub-model discrepancy) : 서브모델 간 동일 어휘에 대한 이진관계 요소의 차이를 제시할 뿐만 아니라(예: 공군 ‘항공작전’과 육/해군의 ‘항공작전’, 해병대 ‘항공작전’) 앞서 제시한 서브모델 간 중복성, 불일치, 순환정의를 식별한다.

5. CMMS-K 검증기능 결과

5.1 구현방법과 테스트데이터

CMMS-K의 온톨로지 데이터는 그래프 구조로 표현할 수 있는데, 개체와 파악은 노드로, 이진관계는 노드 간의 링크로 표현되는 그래프이다. 일관성 검증기능은 이러한 그래프를 대상으로 수행한다.

먼저 테스트데이터를 구축하였으며, 전체 온톨로지 데이터의 일부분으로써, 공군 도메인 부분을 선택했다. 테스트데이터는 전체 826개의 어휘와 1238개의 이진관계를 포함한다. 개념사전은 어휘사전에서 개념으로 분류된 어휘를 따로 목록으로 명세되어 있으며, 개념 별로 속성과 관련 이진관계 등의 내용을 갖는다. 기타 단계의 내용들도 상호 교차검증이 가능한 내용들로 채워져 있다. 테스트데이터는 유형별 오류를 의도적으로 포함한다.

Fig. 7은 테스트데이터를 구성한 방법의 일부를 보여준다. 그림의 표에서 비고 컬럼에 “오류데이터”라고 표시된 부분이 의도적으로 삽입된 오류데이터이며, 이외에도 유형별 오류를 삽입하였다.

2	관계이름 (Relation Concept)	소스 개념 (Source Concept)	목표 개념 (Target Concept)	비고(수정사유)
1148	is-a-part-of	경로	CSAR	
1149	is-a-part-of	고도	CSAR	
1150	is-a-part-of	속도	CSAR	
1151	is-a-part-of	조종사구조	CSAR	
1199	is-a-part-of	A◇◇ 방어사령부	◇◇방어사령부	오류 데이터
1200	is-a-part-of	B◇◇ 방어사령부	◇◇방어사령부	오류 데이터
1201	is-a-part-of	C◇◇ 방어사령부	◇◇방어사령부	오류 데이터
1228	is-a-part-of	A△△△전대	○○사령부	오류 데이터
1229	is-a-part-of	B△△△전대	○○사령부	오류 데이터
1230	is-a-part-of	C△△△전대	○○사령부	오류 데이터
1231	is-a-part-of	D△△△전대	○○사령부	오류 데이터
1237	is-a-part-of	○○사령부	D△△△전대	오류 데이터

Fig. 7. Part of test-data

5.2 구현결과 및 알고리즘 복잡도

단순한 중복성 검증의 경우는, 개별 어휘(1개)마다 다른 전체 n-1 개의 어휘와 비교하는 순차탐색(sequential search)을 활용하며, 따라서 한 어휘에 대한 순차탐색의 복잡도가 O(n)이므로 이 처리를 모든 어휘에서 반복하니 전체 복잡도는 O(n²)이다. Fig. 8은 중복성 검증 결과를 보여준다. 중복성 검사결과와 첫 번째 행에 “89 GPS 유도”의 경우 “729 GPS유도”와 중복된다고 표시되어 있는데, 완전히 일치하지 않고 띄어쓰기에 차이가 있다라도 같은 개념(어휘)으로 보고 중복관단을 한 것이다. 개념 앞의 숫자는 인식번호(ID)이다.



Fig. 8. Result of duplication and missing link check

불일치 검증도 이와 유사하다. Fig. 8의 두 번째 부분(“Checking for concepts if in Glossary...”)을 보면 “34 Air interdiction is not found in glossary”로 “Air interdiction”이 어휘사전에 존재하지 않음을 식별했다. 어휘사전의 어휘가 개념(concept)으로 분류되어 있으면 그 어휘는 개념사전에서 다시 명세되어야 하거나 반대로 개념사전에 개념으로 명세되어 있으면 어휘사전에 개념 유형의 어휘로 등재되어 있어야 한다.

불일치 검증은 특정 개념 또는 어휘가 명세되어야 할 온톨로지정의서의 단계별 양식을 교차 검증한다. 따라서 하나의 검사대상을 각 단계의 양식에 포함된 어휘들을 순차탐색으로 비교하는데, 이 처리가 전체 어휘에 대해 이루어지므로 복잡도는 중복성 검증과 마찬가지로 O(n²)이다.

순환정의 검증은 테스트데이터의 이진관계에 대해 수행된다. Fig. 9는 “is-a-part-of” 이진관계이다. ○○사령부와 ◇◇방어사령부에 소속된 하위부대 간 관계를

보여준다. 이 그림에서 식별할 수 있는 내용은 “○○사령부”와 “D△△△△전대”가 순환정의에 빠져있음을 알 수 있다. 물론 의도적으로 포함시킨 오류이다. “is-a-part-of” 이진관계에 대한 순환정의 검증을 실행하면 Fig. 10과 같은 결과를 얻게 된다.

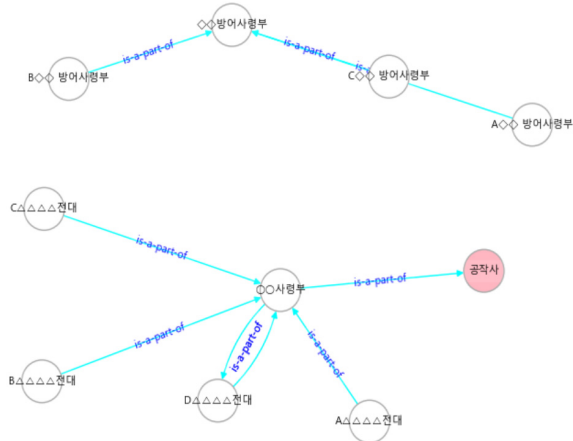


Fig. 9. Binary relation “is-a-part-of”

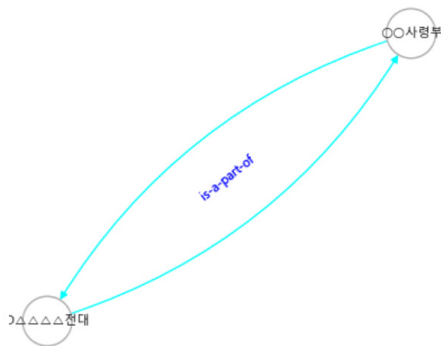


Fig. 10. A circle in “is-a-part-of” relations

하나의 이진관계에 대한 순환정의를 검증하기 위해서는 링크(이진관계 어휘)에 연관된 모든 노드(개념 어휘)를 확인해야 하는데, 깊이우선탐색(depth first search, DFS) 알고리즘을 적용했다. 따라서 복잡도는 일반적인 DFS의 복잡도인 $O(n)$ 이다. 그런데 전체 노드와 링크에 해당되는 어휘(n)에 대해 순환정의를 식별한다면 $O(n^2)$ 이 된다. Fig. 11은 검증기능 SW에서 “a-part-of”와 “is-a” 이진관계에 대한 순환정의를 식별한 결과이다. Fig. 12는 전체 이진관계에 대한 순환정의를 식별한 화면을 보여준다.

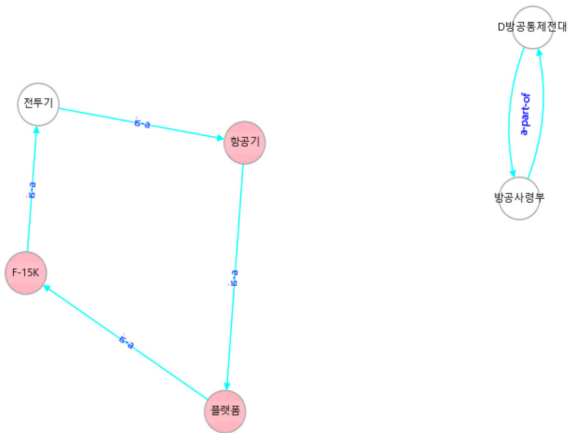


Fig. 11. circles in “a-part-of” and “is-a”

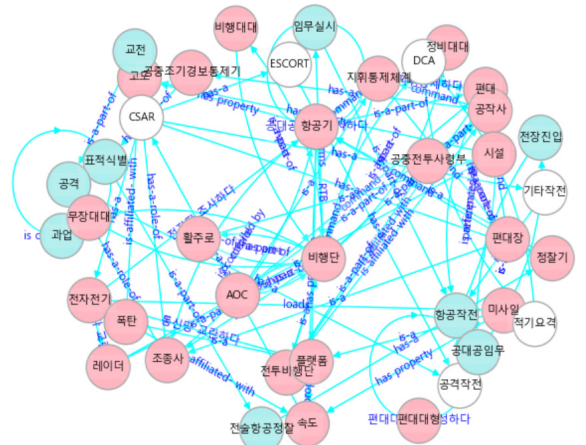


Fig. 12. All circles detected

표기 오류는 띄어쓰기와 사소한 오타에 의한 오류가 있을 수 있다. 예를 들어 띄어쓰기의 경우 “유도미사일”과 “유도 미사일”, 오타의 경우 “해병대”와 “해병대”가 있을 수 있다. 띄어쓰기의 경우는 스페이스를 삭제해서 중복성 검사만 할 수도 있으나 경우에 따라 띄어쓰기가 필요한 경우도 있다. 따라서 유사용어의 경우 문자유사성 식별 알고리즘인 Jaro-Winkler 알고리즘을 활용하였다. 그런데, Jaro-Winkler 알고리즘은 비교적 긴 시간이 필요해서 순차탐색과 조합하여 비교할 경우 복잡도 $O(n^2)$ 으로 n이 증가하면 시간은 n^2 으로 증가하기에 적용이 곤란하다. 따라서 일단 비교대상을 먼저 정렬(복잡도 $O(n \log n)$)한 후 근처에 있는 3개 단어만 비교하는 것(복잡도 $O(3n)$)으로 개선했다. 3개

이상의 경우 비교대상들의 차이가 커서 표기오류 대상이 되지 않으리라 가정한 것이다. 결과적으로 복잡도는 $O(3n + n \log n \rightarrow n \log n)$ 으로 개선되었다.

서브모델간 차이 식별은 각 서브모델에 대해 앞의 유형별 오류 식별을 하고, 두 결과에 대한 단순 1:1 비교를 하기에 복잡도는 $O(n)$ 이 추가되며 결과적으로 단순비교는 무시되고 유형별 오류 복잡도가 적용된다.

실데이터를 검증할 때 기능의 정확도는 다음과 같이 예상된다. 중복성 및 일치성, 순환정의 검증은 두 비교대상 개념의 어휘가 일치하면 식별되므로 모든 오류가 식별될 것이다. 그런데 표기오류의 경우에는 기준 어휘 대비 대상 어휘의 유사수준(Jaro Similarity)으로 식별되기 때문에 비교값 차이가 큰 경우 표기 오류가 아닌 다른 어휘로 판단하여 식별되지 않을 수 있다. 예를 들어 Fig. 13을 보면 표기오류 판단 수준을 엄격(Strict), 중간(Moderate), 여유(Easy) 중 ‘여유’를 선택할 경우, 서로 다른 개념인 ‘전략정찰기’와 ‘전술정찰기’가 ‘여유(Easy)’ 수준에서 오류로 식별된다. 이는 현 CMMS-K 수준에서는 인간이 개입해야 하는 부분도 있다는 것을 알 수 있다.

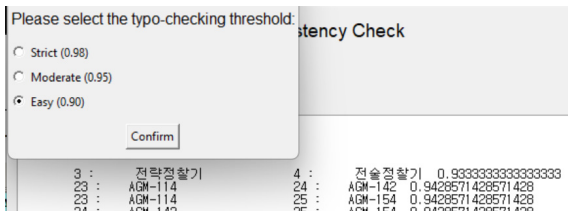


Fig. 13. Typo check result

5.3 검증기능의 함의

CMMS-K의 검증기능은 온톨로지의 품질 향상에 필요하다. 본 연구에서 시도된 검증기능 구현은 검증가능한 부분 중 일부분이며, 전형적인 온톨로지정의서 양식(Methodontology Table)에서 명세된 각종 요소 간 교차검증 가능성을 보여준다. 이 가능성은 다양한 목적의 검증과 추론 기능 개발을 포함한다. CMMS-K 온톨로지의 이러한 특성을 모의모델 개발에 활용할 경우 모의모델 자체의 일관성을 높일 수 있을 것이다.

본 논문에서 제시한 사례는 개체 내용만을 포함하고 있으나 과업에 대한 내용도 사실상 대동소이하여 지면 제한 상 개체 내용만 포함하고 있다. 그러나 구현된 기능은 과업에 대해서도 동일하게 적용할 수 있다. 또한 현 단계의 온톨로지는 구축이 진행 중이며

완결되지 않았다. 향후 제대로 구축된 온톨로지 대상으로 얼마나 일관성이 있는가에 대한 평가를 수행할 예정이다.

6. 결론

해외선진국과 같이 우리도 이러한 온톨로지를 개발하기 위해 한국형 임무공간 개념모델(CMMS-K)을 개발하기 위해 노력하고 있다. 본 논문은 CMMS-K 개념모델에 대한 일관성 검증기능 개발노력을 소개했다. 현재 CMMS-K가 연구단계이지만 연구의 일환으로서 검증기능의 구현 가능성을 입증하였다. CMMS-K는 전형적인 양식을 기반으로 데이터구조를 설정하였기 때문에 데이터 요소별로 교차검증이 가능하였다. 향후 CMMS-K의 실제 데이터에 대해 검증을 수행할 예정이며, 온톨로지가 구축되는 단계, 운용단계, 온톨로지 유지보수 단계에서 일관성 검증기능이 활용되어 CMMS-K의 신뢰성과 활용성을 제고할 것이다.

후 기

이 논문은 국방기술진흥연구소의 지원을 받아 2022년부터 수행 중인 연구임(16-202-210-025, 한국형 임무공간 개념모델 자동화 생성 및 검증 기술).

References

- [1] Morosoff, P., Rudnicki, R., Bryant, J., Farrell, R., & Smith, B., Joint Doctrine Ontology: A Benchmark for Military Information Systems Interoperability : 88ABW-2015-2881, STIDS 2015 Proceedings Page 2, 2015.
- [2] O. Corcho, M. Fernández-López and A. Gómez-Pérez, "Methodologies, tools and languages for building ontologies. Where is their meeting point?," Data & Knowledge Engineering, 46(1), pp. 41-64, 2003.
- [3] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi and P. F. Patel-Schneider, The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, 2003.

- [4] A. Gómez-Pérez, “Ontology Evaluation,” Handbook on ontologies, Springer, pp. 251-273, 2004.
- [5] Andrew Burton-Jones a, Veda C. Storey a, Vijayan Sugumaran b, Punit Ahluwalia, “A semiotic metrics suite for assessing the quality of ontologies,” Data & Knowledge Engineering, 2005.
- [6] Gómez-Pérez, A., Fernández-López, M., Corcho, O., “3.3.5 METHONTOLOGY,” in Ontological Engineering, Springer, pp. 125-142, 2003.
- [7] Shvaiko, P., & Euzenat, J., “Ontology matching: State of the art and future challenges,” IEEE Transactions on Knowledge and Data Engineering, 25(1), pp. 158-176, 2013.
- [8] C. Schlenoff, M. Gruninger, F. Tissot, J. Valois, J. Lubell and J. Lee, “The Process Specification Language(PSL): Overview and Version 1.0 Specification,” NISTIR 6459, National Institute of Standards and Technology, 2000.