

# 지능형 문서변환시스템을 위한 온톨로지구축

임성신\* · 이석용\* · 박남규\*\* · 서창갑\*\*

\*부산대학교, \*\*동명정보대학교

## Development of Ontology for Intelligent Document Transformation System

Sung-shin Lim\* · Seok-yong Lee\* · Nam-kyu Park\*\* · Chang-gab Seo\*\*

\*Pusan National University, \*\*Tongmyoung University of Information Technology

E-mail : {sslim, mislsy}@pusan.ac.kr, {nkpark,gabida}@tit.ac.kr

### 요 약

세계화의 중심에 있는 많은 기업은 특히 수출입업무에서 업무환경의 다변화로 업체와 공공기관에 전달할 데이터가 날로 증가하고 있다. 그래서 이의 효율적인 처리를 위해 많은 기업은 문서변환 시스템을 개발하거나 도입하여 활용하고 있다. 하지만, 기존의 문서변환 시스템은 XML문서변환을 중심으로 개발되고 있으나, 실제 수출입업무처리에서는 XML뿐만 아니라 EDI나 자체적인 파일형식으로 문서를 변환해야 하는 경우도 많다. XML을 중심으로 한 기존연구에서는 수작업에 의한 XML 스키마 연결의 비효율성을 해결하기 위하여 온톨로지를 이용하고 있다. 하지만, 문서변환시스템의 특성상 한번에 완벽하게 온톨로지를 구축할 수가 없다. 그래서 문서변환시스템을 사용하는 사용자에게 온톨로지를 추가, 수정 및 삭제할 수 있는 기능이 보완되어야 한다. 이를 위해 본 연구에서는 지능형 문서변환시스템의 핵심인 온톨로지 구축을 지원하는 저작도구를 문서변환시스템에 연동하고, 이를 이용한 수출입업무 문서변환에 활용할 수 있는 온톨로지를 구축하였다.

### ABSTRACT

The document transformation system is more widely used in order to transform business documents efficiently in diverse organization. In established researches on document transformation systems have been carried mainly focused on XML however, it is not only transformed XML form but also EDI or local form in realistic import and export process. Particularly, in the most completed research relate on document transformation, they used ontology to get rid of non-efficiency in the connection of XML schema by manual. Hence, those researches are lack of features, which are construct and modify the domain ontology automatically and the size wasn't enough to realize itself. In this paper we study development of ontology and basic system, which is critical in intelligent document conversion system. And we develop an ontology with editor can be modified and complemented by users, as well as we make it used in real import and export business process.

### 키워드

문서변환(Document transformation), 온톨로지(Ontology)

### 1. 서 론

수출입업무처리에서 기업의 업무환경 다변화로 업체와 공공기관별로 전달할 데이터양은 날로 증가하고 있다. 이의 효율적인 처리를 위해 UN에서 문서표준기술인 UN/EDIFACT를 제정 발표하였다. 문서표준이 제정 발표되기 전까지는 이에

관한 규약이 없어 기업 내에서도 상호운용에 대한 고려 없이 서로 다른 목적에 의해서 개별적으로 시스템을 개발 구축하여 서식의 형태 및 내용이 조금씩 다르고 표준화되어있지 않아 거래 상대방에서 서류나 전송된 자료를 수정 없이 그대로 사용하지 못하고 이를 재입력 또는 사용자에게 맞게 서식을 고쳐야 하는 문제가 발생하고 있다

[1]. 그리고 도입되었던 여러 애플리케이션이나 시스템 등은 유연하고 빠른 처리를 위하여 프로세스 차원에서 통합되고 연동하여야 할 필요성이 대두하고 있다[2]. 이와 같은 요구를 가장 효과적으로 만족시키는 방법은 범용적으로 사용할 수 있는 문서변환시스템을 이용하는 것이다.

기존의 문서변환과 관련한 연구는 XML 문서의 변환을 중심으로 이루어지고 있는데, 주고 사용자의 수작업에 의한 매핑관계로부터 변환용 XSLT를 생성하고 있다[3, 4]. 이 경우 용량 큰 스키마 처리에는 너무 많은 시간과 비용이 소요된다. 이의 해결을 위해 자동 스키마 매핑에 대한 연구를 수행하고 있다[5, 6, 7, 8]. 자동 스키마 매핑을 위해 온톨로지를 이용하고 있으나 사용자 피드백에 의해 온톨로지를 추가, 수정 및 삭제할 수 있는 기능이 없고 그 규모 또한 실험수준이다.

이를 해결하기 위해 본 논문에서는 지능형 문서변환시스템의 핵심인 온톨로지 구축을 지원하는 저작도구를 문서변환시스템에 연동하고, 이를 이용한 수출입업무 문서변환에 활용할 수 있는 온톨로지를 구축하였다.

## II. 본 론

### 2.1 연구의 범위

일반적으로 우리나라의 수출입 물류프로세스는 수출은 통관·검역, 내륙운송, 선적·기적, 관세환급 순으로 이루어지며, 수입은 하역·하기, 보세운송, 통관·검역, 반출 순으로 이루어지게 된다[9]. 이의 과정에서 발생하는 다양한 문서들의 변환업무는 표준자료의 변환뿐만 아니라 비표준자료를 표준 EDI 및 XML로 변환 또는 Legacy DB에 직접 저장 등 다양하다. 그림1은 개발하고자 하는 시스템의 전체 구성도를 나타낸 것이다.

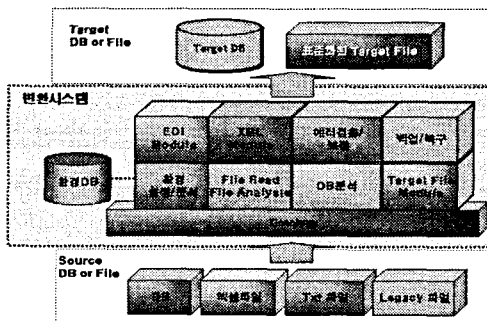


그림 1. 온톨로지기반 문서변환시스템의 구성도

그림1에서 보듯이 전체시스템은 소스파일의 입력, 문서변환, 변환 후 서식생성의 부분으로 구성된다. 본 논문에서는 지능형 문서변환을 위한 온톨로지 구축에 중점을 둘 것이다.

### 2.2 온톨로지

온톨로지(Ontology)는 본래 철학적인 개념으로 "존재의 본질을 연구하는 형이상학의 한 갈래"이나, 자연언어처리에서는 "실세계(혹은 특정 도메인)에 존재하는 모든 개념들과 개념들의 속성, 그리고 개념들이 상호 간 의미상으로 어떻게 연결되어 있는가에 대한 정보를 가지고 있는 지식베이스"로 정의할 수 있다. 그리고 온톨로지는 언어 독립적인 정보만 저장하고 있어서 지식공유와 재사용을 중요시한다는 점과, 개념간 의미관계가 계층관계, 격관계, 동의관계 외의 "has-member, material-of, represent"와 같은 다양한 의미관계도 포함하고 있다는 점에서 시소러스와 구별된다. 그래서 온톨로지는 의미망과 비슷한 원리와 구성을 가진다고 할 수 있다[10]. 그리고 [표1]은 국내의 온톨로지 구축현황을 정리한 것이다[11].

표 1. 국내의 온톨로지 구축현황

구분	내용
국외	Mikrokosmos Ontology
	가도카와(Kadokawa) 시소러스
	EDR 개념사전
	SUMO
	CYC
	WordNet
국내	유로워드넷(EuroWordNet)
	ETRI의 한국어 명사 개념망
	포항공대의 LIP(Language Independent and Practical) 온톨로지
	카이스트 (Kaist Korterm)의 한국어 명사 의미 계층 구조 코어넷(CoreNet)
	울산대(U-WIN)
부산대(KorLex)	
21세기 세종 계획의 전자사전 개발	

기존의 온톨로지는 자연어 처리와 범용적인 목적으로 만들어져 본 연구에서 개발하고자 하는 수출입문서의 변환시스템에 직접적으로 도입하여 사용하기에는 부족함이 많다. 그래서 본 논문에서는 수출입문서의 변환에 사용할 수 있도록 수작업으로 온톨로지를 직접 구축하였다.

### 2.3 온톨로지 저작도구

온톨로지를 이용한 기존의 문서변환시스템의 가장 큰 문제점은 사용자의 요구를 반영한 온톨로지의 추가, 수정 및 삭제를 할 수 없다는 것이다. 그래서 본 연구에서는 이의 해결을 위해 문서변환시스템을 사용하는 사용자가 온톨로지를 보완할 수 있도록 온톨로지 저작도구를 문서변환시스템에 연동했다. 기존의 온톨로지 구축도구들은 단지 온톨로지 편집에 중점을 두었을 뿐 온톨로지 구축자의 프로세스에 따른 여러 고려사항을

염두에 두지 않았다. 개발한 온톨로지 저작도구는 기존의 언어자원(사전, 웹 연동)을 이용하여 사용자가 다양하게 참조할 수 있는 환경을 구축하였다.

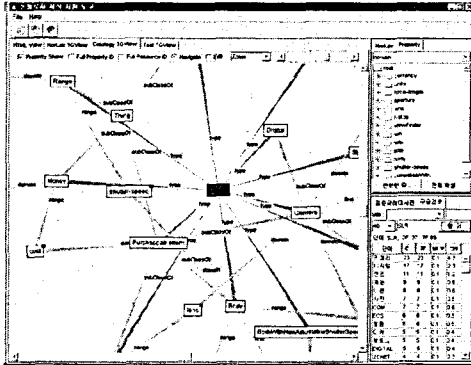


그림 2. 온톨로지 저작도구

#### 2.4 온톨로지 구축과정

수출입 문서변환을 위한 온톨로지의 기본 개념 단위는 워드넷[12]에서와같이 동의어 집합인 synset(set of synonym) 단위로 그림4의 단계를 고려하며 구축하였다[13].

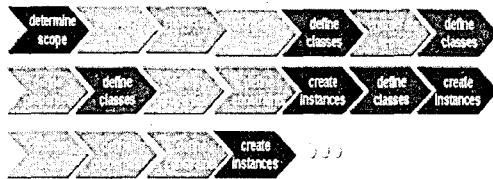


그림 3. 온톨로지의 구축단계

온톨로지는 워드넷, 무역사전, 전자상거래용어사전, 수출입문서 및 물류용어사전을 기반으로 많이 사용되는 용어들에 대한 분석작업을 선행한 후 300여개의 synset을 1,700여개의 단어로 수작업 구축하였다. 하지만, 이 온톨로지는 계속 보완할 것이다.

### III. 결 론

본 논문에서는 지능형 문서변환시스템의 핵심인 온톨로지 구축을 지원하는 저작도구를 문서변환시스템에 연동하고, 이를 이용한 수출입업무 문서변환에 활용할 수 있는 온톨로지를 구축하였다. 하지만, 수작업에 의한 온톨로지의 구축은 성능은 좋지만 구축한 규모에 한계가 있다. 그래서 통계적 처리를 통한 온톨로지의 보완을 수행할

것이다. 그리고 현재의 온톨로지는 수출입문서의 변환을 목적으로 구축하였으나 다른 분야로의 확장을 위해 온톨로지를 보완할 것이다.

#### 참고문헌

- [1] 오재인, "우리나라 물류EDI의 문제점과 활성화 전략: 탐색적 연구", 한국정보전략학회지 제7권 1호, pp. 103-119, 2004
- [2] 송종철, 김창수, 정희경, "스키마 구조 데이터 매핑을 이용한 XML 구조변환 시스템", 정보과학회논문지 컴퓨팅의실제 제10권 제5호, pp. 406-418, 2004. 10
- [3] Eila Kuikka, Paula Leinonen, Martti Penttonen, "Towards Automating of Document Structure Transformations", Proceedings of the 2002 ACM symposium on Document engineering, pp. 103-110, 2002
- [4] Microsoft BizTalk Mapper, <http://www.microsoft.com/biztalk/>
- [5] 이준승, 이경호, "XML 문서의 자동변환을 위한 스키마 매칭 알고리즘", 멀티미디어학회 논문지 제 7권 제9호, pp. 1195-1207, 2004. 9
- [6] S. Bergamaschi, S. Castano, S. De Capitani di Vimercati, S. Montanari, M. Vincini, "An Intelligent Approach to Information Integration", First International Conference on Formal Ontology in Information Systems, pp. 253-267, 1998
- [7] 조정길, 조윤기, 구연설, "구조적 상이성 분석에 기반한 XML 문서 변환 시스템의 설계 및 구현", 정보처리학회논문지 D 제9-D권 제2호, pp. 297-306, 2002. 4
- [8] Stephan Phillippe, Jacob Koheler, "Using XML Technology for the Ontology-Based Semantic Integration of Life Science Databases", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL.8, NO.2. pp. 154-160, 2004. 6
- [9] 김승철, "수출입물류정보 전달 체계 개선을 통한 전자무역 활성화 방안에 관한 연구", 물류학회지 제14권 제2호, pp. 141-164, 2004. 7
- [10] 최호섭, 옥철영, "한국어 의미망 구축과 활용", 한국어학 제17집, pp. 301-329, 2002. 12
- [11] 황순희, "어휘의미망의 이해와 명사 어휘의 미망 구축", 어휘의미망의 이해와 응용 세미나, 2005. 8
- [12] Christiane Fellbaum. "WORDNET: AN ELECTRONIC LEXICAL DATABASE", MIT Press, 1998

[13] Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology", <http://protege.stanford.edu/>