

효율적인 수출입문서 처리를 위한 온톨로지 기반 문서변환시스템 개발

임성신^a 이석용^b 박남규^c 서창갑^d

^a 동명정보대학교 정보기술원
부산 시 남구 용당동 535번지, 608-711
Tel: +82-51-610-8288, Fax: +82-51-610-8289, E-mail: sslim@tit.ac.kr

^b 동명정보대학교 정보기술원
부산 시 남구 용당동 535번지, 608-711
Tel: +82-51-610-8286, Fax: +82-51-610-8289, E-mail: mislisy@tit.ac.kr

^c 동명정보대학교 경영대학 유통경영학과
부산 시 남구 용당동 535번지, 608-711
Tel: +82-51-610-8481, Fax: +82-51-610-8499, E-mail: nkpark@tit.ac.kr

^d 동명정보대학교 경영대학 경영정보학과
부산 시 남구 용당동 535번지, 608-711
Tel: +82-51-610-8443, Fax: +82-51-610-8439, E-mail: gabida@tit.ac.kr

Abstract

수출입업무활동에서 업체와 공공기관별로 전달해야 하는 데이터가 날로 증가하고 있지만, 전달되는 데이터를 보면 기관별 서식이 조금씩 다를 뿐 반복적인 작업의 연속이다. 그래서 이의 효율적인 처리를 위해 많은 기업은 문서변환 시스템을 개발하거나 도입하여 활용하고 있다. 하지만, 기존의 문서변환 시스템은 XML 문서변환을 중심으로 개발되고 있으나, 실제 수출입업무처리에서는 XML 뿐만 아니라 EDI나 자체적인 파일형식으로 문서를 교환해야 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. 이에 본 논문은 수출입 문서변환업무에 사용할 수 있도록 XML 형식뿐만 아니라, EDI 형식이나 자체파일형식으로도 변환할 수 있는 범용적인 온톨로지 기반 문서변환시스템을 구현하였다. 특히 온톨로지의 효과적인 관리를 위하여 문서변환시스템에 온톨로지 저작도구를 연동하여 시스템을 구성하였다.

Keywords:

문서변환; 온톨로지

1. 서론

세계화의 중심에 있는 많은 기업은 특히 수출입업무에서 업무환경의 다변화로 업체와 공공기관에 전달할 데이터가 날로 증가하고 있다. 이의 효율적인 처리를 위해 UN에서 문서표준기술인 UN/EDIFACT를 제정 발표하였다. 그러나 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 표준화되어 있지 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 주 고객 사용자의 수작업에 의한 매칭관계로부터 변환용 XSLT를 생성하고 있다[3, 4]. 이 경우 용량 큰 스키마 처리에는 너무 많은 시간과 비용이 소요된다. 이의

해결을 위해 자동 스키마 매칭에 대한 연구를 수행하고 있다[5, 6, 7, 8]. 자동 스키마 매칭을 위해 온톨로지를 이용하고 있으나 사용자 피드백에 의해 자동으로 도메인 온톨로지를 구축 및 수정할 수 있는 기능이 부족하고 그 규모 또한 실험수준이다.

실제 무역업무처리에서 변환해야 하는 문서는 XML 형식뿐만 아니라 EDI나 자체적인 파일 형식을 이용하는 경우가 많아서, 전달받은 데이터를 내부의 업무처리시스템에 변환하거나 전달해야 하는 형식으로 변환하기 위한 범용적인 변환 시스템이 필요하다. 이에 본 논문은 수출입 문서변환업무에 사용할 수 있도록 XML 형식뿐만 아니라, EDI 형식이나 자체파일형식으로도 변환할 수 있는 범용적인 온톨로지 기반 문서변환시스템을 구현하였다. 특히 온톨로지를 사용자가 추가로 수정 보완할 수 있도록 문서변환시스템에 온톨로지 저작도구를 연동하여 개발하였다.

2. 연구의 범위

일반적으로 우리나라의 수출입 물류프로세서는 수출은 통관·검역, 내륙운송, 선적·기적, 관세환급 순으로 이루어지며, 수입은 하역·하기, 보세운송, 통관·검역, 반출 순으로 이루어지게 된다. 그리고 각 과정에서 발생하는 문서와 관련조직은 Table 1에 정리하였다[9].

Table 1 - 수출입 물류프로세서와 관련조직 및 관련문서

절차	주요업무	관련조직	관련문서	
수출	통관 및 검역	수출신고	화주, 관세사, 관세청	수출신고서, 통관의뢰서, 수출신고필증
		검역	화주, 대행업체, 검역기관	검역신청서, 검역증, 검역결과통지
	운송	운송의뢰 운송	화주, 포워더, 보세운송사	견적의뢰서, 계약서, 운송요청서, 운송계획서, 운송통지서
	선/기적	선적 및 기적예약	해상 화주, 포워더, 선사	운송계약서, 선적요청서, P/L, C/I, L/C사본, 혼재화물적하목록, 적하목록, House B/L
선적 및 기적 적하목록 제출		항공 화주, 포워더, 선사		Booking Request, 운상스케줄, 혼재화물적하

		기적 서류발송	사	목록, Master P/L
	관세 환급	관세환급	화주, 관세사, 관세청	관세환급신청서, 기납증 등 국내거래관련 서류, 환급금결정통지
수입	하역 및 하기	입항 적하목록 신고 실물분류 작업	화주, 포워더, 선사 및 항공사, 해양수산부, 건설교통부, 관세청, 관세사, 하역및조업사	화물적재목록, 화물배정현황표, 입항허가, 화물적하목록, 혼재화물적하목록, 보세운송신고
	보세 운송	운송의뢰	보세운송사	견적의뢰서, 계약서, 운송요청서, 운송계획서
	통관 및 검역	수입신고 검역 보세운송	화주, 관세사, 세관, 검역기관, 보세운송사, 세관, 보세창고	통관의뢰서, 수입신고서, 수입신고필증, 운임지불청구서, 보세운송필증, 보세운송승인서, 화물반입보고
	반출	화물반출	화주, 포워더, 보세창고, CY, 보세운송사	D/O, L/G거래명세, 화물반출보고

Table 1에서 보는 것처럼 수출입과 관련된 업무처리에서 수많은 문서가 요청됨을 알 수 있다. 본 연구에서는 이의 효율적인 처리를 위한 문서변환 시스템의 개발을 목표로 한다.

3. 개발시스템

3.1 시스템 개요

본 논문에서는 표준자료의 변환뿐만 아니라 비표준자료를 표준 EDI 및 XML로 변환 또는 Legacy DB에 직접 저장이 가능한 온톨로지 기반 범용 문서변환시스템을 개발하고자 한다.

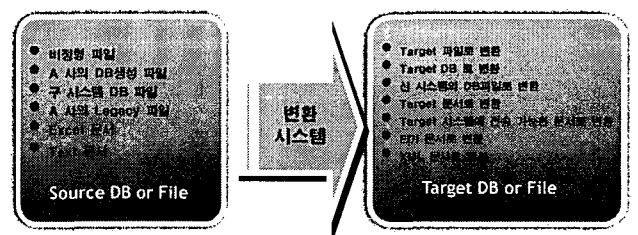


Figure 1 - 연구의 목표

Figure 2는 개발하고자 하는 시스템의 전체 구성도를 나타낸 것이다. 전체적으로 소스파일의 입력, 문서변환, 변환 후 서식생성의 부분으로 구성된다.

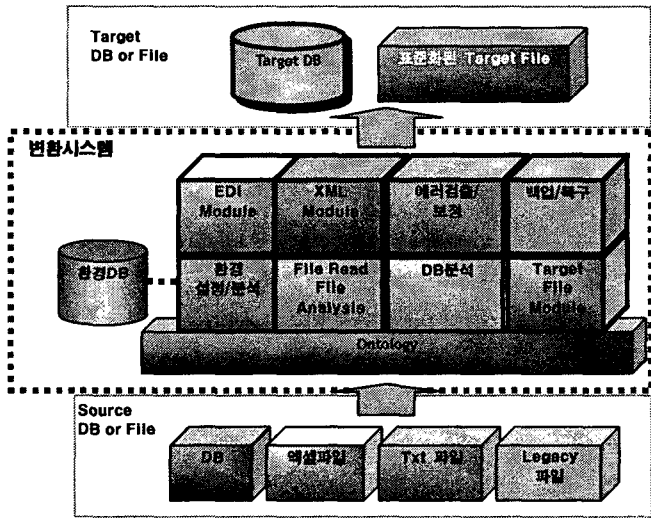


Figure 2 - 전체시스템 구성도

3.2 온톨로지

온톨로지라는 말은 희랍어 ‘ontos(being)’와 ‘logos(word)’에 기인한다. 이는 원래 철학, 특히 형이상학의 한 분야로, 이 세계에 존재(being)하는 것들의 종류, 그 본성과 관계 등에 대한 연구나 학문을 지칭하는 말이다. 이것이 철학에 도입된 것은 19세기의 독일 철학자들에 의해서였는데, 존재론을 자연과학에서 다양하게 언급하는 존재들과 구분하기 위해서였다. 철학적 관점에서 본다면 온톨로지는 세상의 어떤 관점을 설명하는 분류체계를 제공하는 것이라 할 수 있다.

이 용어가 전산학 문헌에서 처음으로 등장한 것은 1967년 S.H. Mealy에 의해서라고 여겨진다. 이후 이 온톨로지는 지난 수십년간 특히 인공지능 분야서 지식표현(knowledge representation), 지식공학 관점에서 다양하게 사용되고 있다. 이런 관점에서 가장 널리 알려진 정의는 Gruber의, “온톨로지란 공유된 개념화(shared conceptualization)의 형식적이고 분명한 명세”라는 정의다. 여기서 개념화란 어떤 목적으로 표현하고자 하는 대상을 추상화하고 단순화시킨 것이다. 형식적이란 규정된 용어들과 그들 사이의 관계를 컴퓨터가 이해할 수 있는 방법으로 표시하는 것이다.

한편 다른 지식공학적 관점에서, “온톨로지란 어떤 특정 영역에서 존재하거나 존재할 수 있는 대상들의 범주(category)에 관한 연구”라고 정의된다. 따라서 이런 연구의 결과인 온톨로지는 가능한 세계의 대상의 목록(catalogue)이라고 할 수 있다. 이런 관점에서 본다면 일반적으로 온톨로지는 대상을 범주화하여 명칭을 붙이거나 개념화된 구조를

충칭하는 의미로 사용되며 각각의 범주명은 개념이나 범주로 언급된다.

그러나 온톨로지에 관한 어떤 보편적인 정의도 없다는 점이 지적되어야 한다. 그 이유중의 하나는 앞서서도 잠시 언급한 대로 그 사용의 광범위함 때문이다. 온톨로지의 개념은 인간과 컴퓨터 시스템 사이의 의사소통을 위해서 적용되기도 하고, 지식의 조직과 재사용(reuse) 관점에서, 그리고 추론(inference)의 관점에서 적용되는 등 다양한 양상을 보인다[10].

Table 2는 국내의 온톨로지 구축현황을 정리한 것이다[11].

Table 2 - 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.

온톨로지 저작도구

온톨로지를 이용한 기존의 문서변환시스템의 가장 큰 문제점은 사용자의 요구를 반영하여 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 연동하였다.

Table 3 - 온톨로지 저작도구 비교

	KAON	OilEd	OntoEdit	Ontolingua	Protege2000
Developers	AIFB-FZI(Univer sity of Karlsruhe)	IMG(University of Manchester)	Ontoprise	KSL(Stanford University)	SMI(Stanford University)
Availability	Open source	Open source	Freeware	Free access	Open source
Extensibility	Application server	Backends	Plug-ins	--	Plug-ins
Ontology storage	Files, DBMS	Files	Files	Files	Files, DBMS
Programming language	Java	Java	Java	Lisp	Java
KR formalism	Frames	DL	Frames+FOL	Frames+FOL	Frames+FOL
Underlying KR language	KAON	DAML+OIL	OXML	Ontolingua	--
Front-end	Java Swing	Java Swing	Java Swing	HTML	Java Swing
Graphical concept taxonomy edition	Yes	No	Yes	No	Yes
Formal axiom editor	No	Yes	No	No	Yes
Collaborative edition	Yes	No	Yes	Yes	No
Inference engine	--	FaCT(built-in) RACER DIG compliant	OntoBroker	JTP	PAL(built-in) JESS, FaCT Prolog, FLORA Algonon
Consistency checking	Yes	Yes	Yes	No	Yes
Automatic classification	No	Yes	No	No	No

Table 3에 기존의 온톨로지 저작도구를 비교하였다[12]. 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.

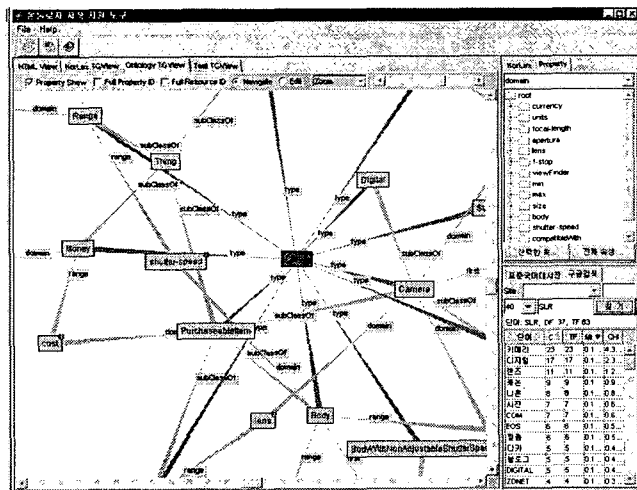


Figure 3 - 온톨로지 저작도구

온톨로지 구축 고려사항

온톨로지의 구축단계를 간략하게 설명하면, 특정한 목적과 영역을 고려한 다음, 개념을 자동·반자동 추출하거나 어휘 사전을 확보하여, 확보된 개념들을 정의하고 조직화해야 한다. 조직화한다는 것은 개념들 간의 관계를 설정함과 동시에 개념이 가지는 특수한 속성을 추출하여 체계화시키는 것을 의미하는 것으로서, 이론적 체계와 더불어 실질적인 구축 원리를 마련해야 한다. 다음 단계로 온톨로지를 표현할 웹 온톨로지 언어나 기계 가독형 표현 언어(machine readable representation language)를 설정하여 형식화하고 실질적으로 구현하여야 하며, 구축중인 온톨로지와의 다른 온톨로지와의 통합 문제와 기존에 존재하는 많은 자원(resources)을 어떻게 활용할 것인가를 모색해야 한다. 그리고 마지막으로 구축된 온톨로지를 대상 애플리케이션(target application)에서 실험하거나 사용패턴(usage patterns)을 분석하여 평가해야 한다. 또한 이러한 평가 결과를 바탕으로 유지·보수를 해야 하며, 조금 더 발전적인 온톨로지로 개선해야 한다.

이러한 일련의 구축 단계에서 가장 어려운 문제는

바로 온톨로지를 실질적으로 구축하는 이론적 체계와 원리가 아직까지 마련되지 않았다는 것이다. 기존의 구축 사례들을 살펴보면, 온톨로지의 실질적인 구축 측면보다도 기존의 시소러스나 의미망, 분류체계 등을 이용한 온톨로지 구축이나 기존에 구축된 온톨로지를 이용한 애플리케이션 개발이 대부분을 차지한다. 이것은 온톨로지 구축에 있어서 국내외적으로 가지는 공통적인 문제라 할 수 있다. 이를 위해서는 WordNet, UMLS와 같이 관련 학문에 대한 이론 습득과 더불어 자연언어처리 기법의 활용을 통한 언어 습득 및 이해 처리 등과 같은 부수적인 연구가 뒤따라야 할 것이다. 이외에 온톨로지 구축에 고려해야 할 사항을 정리하면 다음과 같다[13].

- 온톨로지 구축 대상자 선정
- 온톨로지 구축 방법(자동/반자동/수동)
- 응용분야별 온톨로지 구축 원리 파악
- 기존에 구축된 시소러스, 분류체계 활용 방법
- 범용적/국제적 수준의 온톨로지 구축
- 온톨로지 저작 도구 개발
- 온톨로지 구축 관련 표준방식 연구
- 시맨틱 마크업 언어 표준안 개발

온톨로지 구축과정

수출입 문서변환을 위한 온톨로지의 기본 개념단위는 워드넷[14]에서와같이 동의어 집합인 synset(set of synonym) 단위로 그림4의 단계를 고려하며 구축하였다[15].

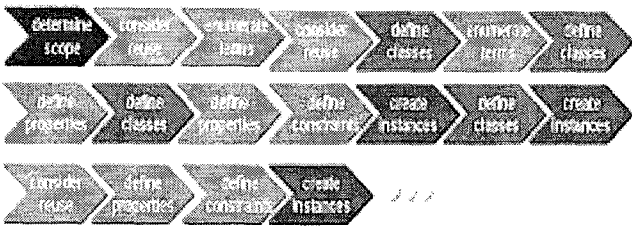


Figure 4 - 온톨로지의 구축단계

구축한 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.

3.3 XML 변환 모듈

XML의 처리 모듈은 Source 문서인 XML문서를 읽어 Temp Table에 임시로 저장한 후 최종 Target으로 생성하거나 Target XML 문서로 생성하기 위해 XML Library 정보를 읽어 XML Library

정보에 맞게 XML 문서를 생성시키는 모듈이다. 업체별 업무에 대한 XML 형태를 협의한 후 업무별 XML Library를 생성하여 서버에 저장한다. 그리고 XML Library 파일을 토대로 변환 및 검정한다.

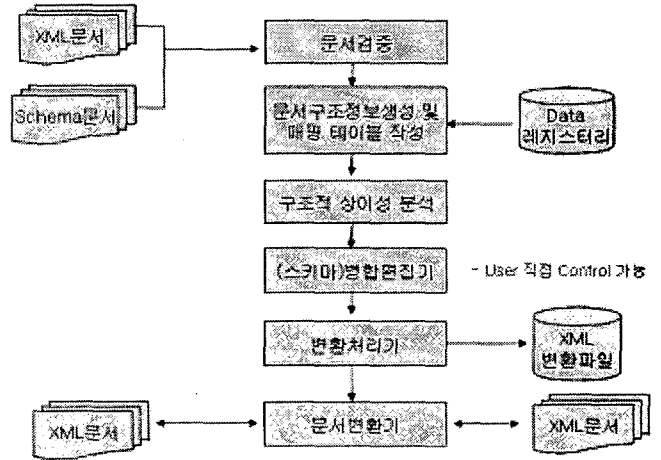


Figure 5 - XML 변환모듈

3.4 EDI 변환 모듈

EDI변환 모듈은 Source 문서인 EDI문서를 읽어 Temp Table에 임시로 저장한 후 최종 Target으로 생성하거나 Target EDI 문서로 생성하기 위해 Flat File를 읽어서 Flat File에 맞게 EDI 문서를 생성하는 모듈이다. 특히 EDI 문서는 업체 간의 협약에 의해 약속된 형태로 file을 생성하여 보낸다. 작업과정상 EDI 파일 정보를 담고 있는 Flat File을 먼저 생성하여 변환 및 검정한다.

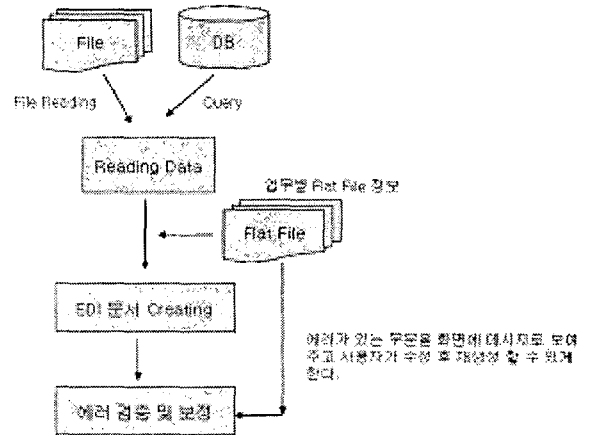


Figure 6 - EDI 모듈

3.5 비정형 파일의 목적파일 변환

비정형 파일을 목적파일로 변환하는 모듈은 Source가 File 또는 EDI문서 또는 XML 문서든 간에 최종 목적물인 File로 생성하기 위해 환경정보를 읽어서 최종 목적물의 형태에 맞게 File를 생성하는 모듈이다. Source는 최종 Target으로 생성하기 위해 중간단계인 Temp Table에 저장한다. 그리고 Temp

Table에서 최종 목적물인 File, EDI 문서, XML 문서로 생성한다.

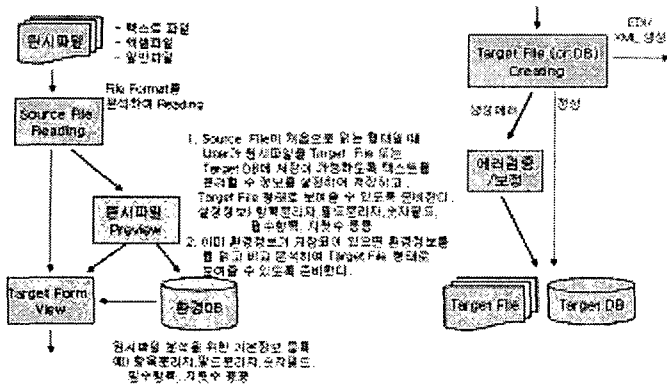


Figure 7 - Source File을 Target File 또는 Target DB에 저장하는 과정에 대한 설명

4. 결론 및 향후연구

본 논문은 수출입 문서변환업무에 사용할 수 있도록 XML형식뿐만 아니라, EDI형식이나 자체파일형식으로도 변환할 수 있는 범용적인 온톨로지 기반 문서변환시스템을 구현하였다.

하지만, 수작업에 의한 온톨로지의 경우 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 추가로 구축할 것이다.

References

[1] 오재인. (2004). “우리나라 물류EDI의 문제점과 활성화 전략: 탐색적 연구”, *한국정보전략학회지* 제7권 1호, pp. 103-119.

[2] 송종철, 김창수, 정희경. (2004). “스키마 구조 데이터 매핑을 이용한 XML 구조변환 시스템”, *정보과학회논문지 컴퓨팅의실제* 제10권 제5호, pp. 406-418.

[3] Eila Kuikka, Paula Leinonen, Martti Penttonen. (2002). “Towards Automating of Document Structure Transformations”, *Proceedings of the 2002 ACM symposium on Document engineering*, pp. 103-110.

[4] Microsoft BizTalk Mapper, <http://www.microsoft.com/biztalk/>

[5] 이준승, 이경호. (2004). “XML 문서의 자동변환을 위한 스키마 매칭 알고리즘”, *멀티미디어학회 논문지* 제 7권 제9호, pp. 1195-1207.

[6] S. Bergamaschi, S. Castano, S. De Capitani di Vimercati, S. Montanari, M. Vincini. (1998). “An Intelligent Approach to Information Integration”, *First*

International Conference on Formal Ontology in Information Systems, pp. 253-267.

[7] 조정길, 조윤기, 구연설. (2002). “구조적 상이성 분석에 기반한 XML 문서 변환 시스템의 설계 및 구현”, *정보처리학회논문지 D* 제9-D권 제2호, pp. 297-306.

[8] Stephan Philippie, Jacob Koheler. (2004) “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.

[9] 김승철. (2004). “수출입물류정보 전달 체계 개선을 통한 전자무역 활성화 방안에 관한 연구”, *물류학회지* 제14권 제2호, pp. 141-164.

[10] 신호필. (2004). “지식기반(Knowledge Base)으로서의 온톨로지(Ontology)와 시맨틱 웹(Semantic Web)”, *정보처리학회지* 제 11권 제2호, pp. 1-12.

[11] 황순희. (2005). “어휘의미망의 이해와 명사 어휘의미망 구축”, *어휘의미망의 이해와 응용 세미나*.

[12] Asuncion Gomez-Perez, Mariono Fernandez-Lopez and Oscar Corcho. (2004). “Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web”, Springer

[13] 최호섭, 옥철영. (2004). “정보검색 시스템과 온톨로지”, *정보과학회지* 제 22권 제4호, pp. 62-71

[14] Christiane Fellbaum. (1998). “WORDNET: AN ELECTRONIC LEXICAL DATABASE”, MIT Press.

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